NONSMOOTH TRUST-REGION ALGORITHM WITH APPLICATIONS TO ROBUST STABILITY OF UNCERTAIN SYSTEMS

Pierre Apkarian[†], Dominikus Noll[∗], Laleh Ravanbod[∗]

ABSTRACT. We propose a bundle trust-region algorithm to minimize locally Lipschitz functions which are potentially nonsmooth and nonconvex. We prove global convergence of our method and show by way of an example that the classical convergence argument in trust-region methods based on the Cauchy point fails in the nonsmooth setting. Our method is tested experimentally on three problems in automatic control.

Keywords. Bundle · cutting plane · trust-region · Cauchy point · global convergence · parametric robustness · distance to instability · worst-case H_∞-norm

1. INTRODUCTION

We consider optimization problems of the form

$$
\begin{array}{ll}
\text{minimize} & f(x) \\
\text{subject to} & x \in C
\end{array}
\tag{1}
$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz, but possibly nonsmooth and nonconvex, and where C is a simply structured closed convex constraint set. We develop a bundle trust-region algorithm for (1), which uses nonconvex cutting planes in tandem with a suitable trust-region management to assure global convergence. The trust-region management is to be considered as an alternative to proximity control, which is the usual policy in bundle methods. Trust-regions allow a tighter control on the step-size, and give a larger choice of norms, whereas bundling is fused on the use of the Euclidean norm. Our experimental part demonstrates how these features may be exploited algorithmically.

Algorithms where bundle and trust-region elements are combined are rather sparse in the literature. For convex objectives Ruszcyński [38] presents a bundle trust-region method, which can be extended to composite convex functions. An early contribution where bundling and trust-regions are combined is [42, 43], and this is also used in versions of the BT-code [46]. Fuduli *et al.* [19] use DC-functions to form a non-standard trust-region, which they also use in tandem with cutting planes. A feature which these methods share with nonconvex bundle methods like Sagastizábel and Hare [39, 40] or [33] is that the objective is approximated by a simply structured, often polyhedral, working model, which is updated iteratively by adding cutting planes at unsuccessful trial steps. Our main Theorem 1 analyses the interaction of this mechanism with the trust-region management, and assures global convergence under realistic hypotheses.

The trust-region strategy is well-understood in smooth optimization, where global convergence is proved by exploiting properties of the Cauchy point, as pioneered in Powell [35]. For the present work it is therefore of the essence to realize that the Cauchy point fails in the nonsmooth setting. This happens even for polyhedral convex functions, the simplest possible case, as we demonstrate by way of a counterexample. This explains why the convergence proof has to be organized along different lines.

[†]Control System Department, ONERA, Toulouse, France.
[∗]Institut de Mathématiques, Université de Toulouse, France.

The question is then whether there are more restrictive classes of nonsmooth functions, where the Cauchy point can be salvaged. In response we show that the classical trust-region strategy with Cauchy point is still valid for upper C^1-functions, and at least partially, for functions having a strict standard model. It turns out that several problems in control and in contact mechanics are in this class, which justifies the disquisition. Nonetheless, the class of functions where the Cauchy point works remains exceptional in the nonsmooth framework, which is corroborated by the fact that it does not include nonsmooth convex functions.

A strong incentive for the present work comes indeed from applications in automatic control. In the experimental part we will apply our novel bundle trust-region method to compute locally optimal solutions to three NP-hard problems in the theory of systems with uncertain parameters. This includes (i) computing the worst-case H_∞-norm of a system over a given uncertain parameter range, (ii) checking robust stability of an uncertain system over a given parameter range, and (iii) computing the distance to instability of a nominally stable system with uncertain parameters. In these applications the versatility of the bundle trust-region approach with regard to the choice of the norm is exploited.

Nonsmooth trust-region methods which do *not* include the possibility of bundling are more common, see for instance Dennis *et al.* [17], where the authors present an axiomatic approach, and [13, Chap. 11], where that idea is further expanded. A recent trust-region method for DC-functions is [26].

The structure of the paper is as follows. The algorithm is developed in section 2, and its global convergence is proved in section 3. Applications of the model approach are discussed in section 5, where we also discuss failure of the Cauchy point. Numerical experiments with three problems in automatic control are presented in section 6.

Notation

For nonsmooth optimization we follow [12]. The Clarke directional derivative of f is $f^\circ(x, d)$, its Clarke subdifferential $\partial f(x)$. For a function ϕ of two variables $\partial_1 \phi$ denotes the Clarke subdifferential with respect to the first variable. For symmetric matrices $M \preceq 0$ means negative semidefinite. For linear system theory see [45].

2. Presentation of the algorithm

In this chapter we derive our trust-region algorithm to solve program (1) and discuss its building blocks.

2.1. Working model. We start by explaining how a local approximation of f in the neighborhood of the current serious iterate x, called the *working model* of f, is generated iteratively. We recall the notion of a first-order model of f introduced in [33].

Definition 1. A function $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called a *first-order model* of f on a set Ω if $\phi(\cdot, x)$ is convex for every $x \in \Omega$, and the following properties are satisfied:

(M_1) $\phi(x, x) = f(x)$, and $\partial_1 \phi(x, x) \subset \partial f(x)$.
(M_2) If $y_k \to x$, then there exist $\epsilon_k \to 0^+$ such that $f(y_k) \leq \phi(y_k, x) + \epsilon_k \|y_k - x\|$.
(M_3) If $x_k \to x$, $y_k \to y$, then $\limsup_{k \to \infty} \phi(y_k, x_k) \leq \phi(y, x)$. $\qquad \square$

We may think of $\phi(\cdot, x)$ as a non-smooth first-order Taylor expansion of f at x. Every locally Lipschitz function has indeed a first-order model $\phi^\sharp$, which we call the *standard model*, defined as

$$\phi^\sharp(y, x) = f(x) + f^\circ(x, y - x).$$

Here $f^\circ(x, d)$ is the Clarke directional derivative of f at x in direction d. Following [33], a first-order model $\phi(\cdot, x)$ is called *strict* at $x \in \Omega$ if the following strict version of (M_2) is satisfied:

$(\widetilde{M_2})$ Whenever $y_k \to x$, $x_k \to x$, there exist $\epsilon_k \to 0^+$ such that $f(y_k) \le \phi(y_k, x_k) + \epsilon_k \|y_k - x_k\|$.

Remark 1. Axiom (M_2) corresponds to the one-sided Taylor type estimate $f(y) \le \phi(y, x) + \mathrm{o}(\|y - x\|)$ as $y \to x$. In contrast, axiom $(\widetilde{M_2})$ means $f(y) \le \phi(y, x) + \mathrm{o}(\|y - x\|)$ as $\|y - x\| \to 0$ uniformly on bounded sets. This is analogous to the difference between differentiability and strict differentiability, hence the nomenclature of a strict model.

Remark 2. Note that the standard model $\phi^\sharp$ of f is not always strict [31]. A strict first-order model ϕ is for instance obtained for composite functions $f = h \circ F$ with h convex and F of class C^1, if one defines

$$\phi(y, x) = h\left(F(x) + F'(x)(y - x)\right),$$

where $F'(x)$ is the differential of the mapping F at x. The use of a natural model of this form covers for instance approaches like Powell [35], or Ruszczyński [38], where composite functions are discussed.

Observe that every convex f is its own strict model $\phi(y, x) = f(y)$ in the sense of definition 1. As a consequence, our algorithmic framework contains the convex cutting plane trust-region method [38] as a special case.

Remark 3. It follows from the previous remark that a function f may have several first-order models. Every model ϕ leads to a different algorithm for (1).

We continue to consider x as the current serious iterate of our algorithm to be designed, and we consider z, a trial point near x, which is a candidate to become the next serious iterate x^+. The way trial points are generated will be explained in Section 2.2.

Definition 2. Let x be the current serious iterate and z a trial step. Let g be a subgradient of $\phi(\cdot, x)$ at z, for short, $g \in \partial_1 \phi(z, x)$. Then the affine function $m(\cdot, x) = \phi(z, x) + g^\top(\cdot - z)$ is called a *cutting plane* of f at serious iterate x and trial step z. $\qquad\square$

We may always represent a cutting plane at serious iterate x in the form

$$m(\cdot, x) = a + g^\top(\cdot - x),$$

where $a = m(x, x) = \phi(z, x) + g^\top(x - z) \le f(x)$ and $g \in \partial_1 \phi(z, x)$. We say that the pair (a, g) represents the cutting plane $m(\cdot, x)$.

We also allow cutting planes $m_0(\cdot, x)$ at serious iterate x with trial step $z = x$. We refer to these as *exactness planes* of f at serious iterate x, because $m_0(x, x) = f(x)$. Every (a, g) representing an exactness plane is of the form $(f(x), g_0)$ with $g_0 \in \partial f(x)$.

Remark 4. For the standard model $\phi^\sharp$ a cutting plane for trial step z at serious iterate x has the very specific form $m^\sharp(\cdot, x) = f(x) + g_z^\top(\cdot - x)$, where $g_z \in \partial f(x)$ attains the maximum $f^\circ(x, z - x) = g_z^\top(z - x)$. Here every cutting plane $m^\sharp(\cdot, x)$ is also an exactness plane, a fact which will no longer be true for other models. If f is strictly differentiable at x, then there is only one cutting plane $m^\sharp(\cdot, x) = f(x) + \nabla f(x)^\top(\cdot - x)$, the first-order Taylor polynomial.

Definition 3. Let $\mathcal{G}_k$ be a set of pairs (a, g) all representing cutting planes of f at trial steps around the serious iterate x. Suppose $\mathcal{G}_k$ contains at least one exactness plane at x. Then $\phi_k(\cdot, x) = \max_{(a,g) \in \mathcal{G}_k} a + g^\top(\cdot - x)$ is called a *working model* of f at x. $\qquad\square$

Remark 5. We index working models ϕ_k by the inner loop counter k to highlight that they are updated in the inner loop by adding tangent planes of the ideal model ϕ at the null steps y^k.

Usually the ϕ_k are rough polyhedral approximation of ϕ, but we do not exclude cases where the ϕ_k are generated by infinite sets $\mathcal{G}_k$. This is for instance the case in the spectral bundle method [20, 21, 22], see also [7], which we discuss this in 5.3.

Remark 6. Note that even the choice $\phi_k = \phi$ is allowed in definition 3 and in algorithm 1. This corresponds to $\mathcal{G} = \{(a, g) : g \in \partial f(z), a = \phi(z, x) + g^\top (x - z)\}$, which is the largest possible set of cuts, or the set of all cuts obtained from ϕ. We discuss this case in section 5.1. If $\phi^\sharp$ is used, then the corresponding working models are denoted $\phi_k^\sharp$. Their case is analyzed in section 5.4.

The properties of a working model may be summarized as follows

Proposition 1. *Let $\phi_k(\cdot, x)$ be a working model of f at x built from $\mathcal{G}_k$ and based on the ideal model ϕ. Then*

(i) *$\phi_k(\cdot, x) \leq \phi(\cdot, x)$.*
(ii) *$\phi_k(x, x) = \phi(x, x) = f(x)$.*
(iii) *$\partial_1 \phi_k(x, x) \subset \partial_1 \phi(x, x) \subset \partial f(x)$.*
(iv) *If $(a, g) \in \mathcal{G}_k$ contributes to ϕ_k and stems from the trial step z at serious iterate x, then $\phi_k(z, x) = \phi(z, x)$.*

Proof. By construction ϕ_k is a maximum of affine minorants of ϕ, which proves (i). Since at least one plane in $\mathcal{G}_k$ is of the form $m_0(\cdot, x) = \phi(x, x) + g^\top(\cdot - x)$ with $g \in \partial_1 \phi(x, x)$, we have $\phi_1(x, x) \geq m_0(x, x) = \phi(x, x) = f(x)$, which proves (ii). To prove (iii), observe that since $\phi_k(\cdot, x)$ is convex, every $g \in \partial_1 \phi_k(x, x)$ gives an affine minorant $m(\cdot, x) = \phi_k(x, x) + g^\top(\cdot - x)$ of $\phi_k(\cdot, x)$. Then $m(\cdot, x) \leq \phi(\cdot, x)$ with equality at x. By convexity $g \in \partial_1 \phi(x, x)$, and by axiom (M_1) we have $g \in \partial f(x)$. As for (iv), observe that every cutting plane $m(\cdot, x)$ at z satisfies $m(z, x) = \phi(z, x)$, hence also $\phi_k(z, x) = \phi(z, x)$. □

2.2. Tangent program. In this section we discuss how trial steps are generated. Given the current working model $\phi_k(\cdot, x) = \max\{a + g^\top(\cdot - x) : (a, g) \in \mathcal{G}_k\}$, and the current trust-region radius R_k, the *tangent program* is the following convex optimization problem

(2)
$$\begin{array}{ll} \text{minimize} & \phi_k(y, x) \\ \text{subject to} & y \in C \\ & \|y - x\| \leq R_k \end{array}$$

where $\|\cdot\|$ could be any norm on $\mathbb{R}^n$. Let y^k be an optimal solution of (2). By the necessary optimality condition there exists a subgradient $g_k \in \partial\left(\phi_k(\cdot, x) + i_C\right)(y^k)$ and a vector v_k in the normal cone to $B(x, R_k)$ at $y^k \in B(x, R_k)$ such that $0 = g_k + v_k$, where i_C is the indicator function of C. We call g_k the *aggregate subgradient* at y^k. This terminology stems from the classical bundle method, when a polyhedral working model is used, see Ruszczyński [38], Kiwiel [24].

Solutions y^k of (2) are candidates to become the next serious iterate x^+. For practical reasons we now enlarge the set of possible candidates. Fix $0 < \theta \ll 1$ and $M \geq 1$, then every $z^k \in C \cap B(x, M\|x - y^k\|)$ satisfying

(3)
$$f(x) - \phi_k(z^k, x) \geq \theta\left(f(x) - \phi_k(y^k, x)\right)$$

is called a *trial step*. Note that y^k itself is of course a trial step, because $f(x) \geq \phi_k(y^k, x)$ by the definition of the tangent program. But due to $\theta \in (0, 1)$, there exists an entire neighborhood U of y^k such that every $z^k \in U \cap C$ is a trial step.

Remark 7. The role of y^k here is not unlike that of the Cauchy point in classical trust-region methods. Suppose we use a standard working model $\phi_k^\sharp$ and f is strictly differentiable at x. Then $\phi_k^\sharp(\cdot, x) = \phi^\sharp(\cdot, x) = f(x) + \nabla f(x)^\top(\cdot - x)$. In the unconstrained case $C = \mathbb{R}^n$ the solution y^k has then the explicit form $y^k = x - R_k \frac{\nabla f(x)}{\|\nabla f(x)\|}$, which is indeed the Cauchy point as considered in [41], see also [38, (5.108)]. Condition (3) then takes the familiar form $f(x) - \phi_k^\sharp(z^k, x) \geq \sigma\|\nabla f(x)\|R_k$, see [38, (5.110)].

2.3. Acceptance test. In order to decide whether a trial step z^k will become the next serious iterate x^+, we compute the test quotient

$$(4) \qquad \rho_k = \frac{f(x) - f(z^k)}{f(x) - \phi_k(z^k, x)},$$

which compares as usual actual progress and model predicted progress. For a fixed parameter $0 < \gamma < 1$, the decision is as follows. If $\rho_k \geq \gamma$, then the trial step z^k is accepted as the new iterate $x^+ = z^k$, and we call this a *serious step*. On the other hand, if $\rho_k < \gamma$, then z^k is rejected and referred to as a *null step*. In that case we compute a cutting plane $m_k(\cdot, x)$ at z^k, and add it to the new set $\mathcal{G}_{k+1}$ in order to improve our working model. In other words, a pair (a_k, g_k) is added, where $g_k \in \partial_1 \phi(z^k, x)$ and $a_k = \phi(z^k, x) + g_k^\top(x - z^k)$.

Remark 8. Adding one cutting plane at the null step z^k is mandatory, but we may at leisure add several other tangent planes of $\phi(\cdot, x)$ to further improve the working model. A case of practical importance, where the ϕ_k are generated by infinite sets $\mathcal{G}_k$ of cuts, is presented in section 5.3.

Remark 9. In most applications ϕ_k is a polyhedral convex function. If C is also polyhedral, then it is attractive to choose a polyhedral trust-region norm $\|\cdot\|$, because this makes (2) a linear program.

Remark 10. For polyhedral ϕ_k one can limit the size of the sets $\mathcal{G}_k$. Consider for simplicity $C = \mathbb{R}^n$, then the tangent program (2) is $p = \min\{t : a_i + g_i^\top(y - x) - t \leq 0, i = 0, \ldots, k, \|z - x\| \leq R_k\}$. Its dual is $d = \max\{\sum_{i=1}^k \lambda_i a_i - R_k\|\sum_{i=1}^k \lambda_i g_i\| : \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1\}$. By Carathéodory's theorem we can select a subset $\{(a_0, g_0), \ldots, (a_n, g_n)\}$ of $\mathcal{G}_k$ of size at most $n + 1$ with the same convex hull as $\mathcal{G}_k$, so it is always possible to limit $|\mathcal{G}_k| \leq n + 1$. This estimate is pessimistic. An efficient but heuristic method is to remove from $\mathcal{G}_k$ a certain number of cuts which were not active at the last z^k. In the bundle method with proximity control, Kiwiel's aggregate subgradient [24] allows a rigorous theoretical limit of $|\mathcal{G}_k| \leq 3$, even though in practice one keeps more cuts in the $\mathcal{G}_k$. It is not known whether Kiwiel's argument can be extended to the trust-region case, and the only known bound is $n + 1$, see also [38, Ch. 7.5] for a discussion.

2.4. Nonsmooth solver. We are now ready to present our algorithm for program (1). See Algorithm 1 next page.

3. Convergence

In this chapter we analyze the convergence properties of the main algorithm.

3.1. Convergence of the inner loop. In this section we prove finiteness of the inner loop with counter k. Since the outer loop counter j is fixed, we simplify notation and write $x = x^j$ for the current serious iterate, and $x^+ = x^{j+1}$ for the next serious iterate, which is the result of the inner loop.

Algorithm 1. Nonsmooth trust-region method

Parameters: $0 < \gamma < \widetilde{\gamma} < 1, 0 < \gamma < \Gamma \leq 1, 0 < \theta \ll 1, M \geq 1$.

▷ **Step 1** (`Initialize outer loop`). Choose initial iterate $x^1 \in C$. Initialize memory trust-region radius as $R_1^\sharp > 0$. Put $j = 1$.

◇ **Step 2** (`Stopping test`). At outer loop counter j, stop if x^j is a critical point of (1). Otherwise, goto inner loop.

▷ **Step 3** (`Initialize inner loop`). Put inner loop counter $k = 1$ and initialize trust-region radius as $R_1 = R_j^\sharp$. Build initial working model $\phi_1(\cdot, x^j)$ based on $\mathcal{G}_1$, where at least $(f(x^j), g_{0j}) \in \mathcal{G}_1$ for some where $g_{0j} \in \partial f(x^j)$. Possibly enrich $\mathcal{G}_1$ by recycling some of the planes from the previous serious step.

▷ **Step 4** (`Trial step generation`). At inner loop counter k find solution y^k of the tangent program

$$\begin{aligned}
\text{minimize} \quad & \phi_k(y, x^j) \\
\text{subject to} \quad & y \in C \\
& \|y - x^j\| \leq R_k
\end{aligned}$$

Then compute any trial step $z^k \in C \cap B(x^j, M\|x^j - y^k\|)$ satisfying $f(x^j) - \phi_k(z^k, x^j) \geq \theta \left(f(x^j) - \phi_k(y^k, x^j) \right)$.

◇ **Step 5** (`Acceptance test`). If

$$\rho_k = \frac{f(x^j) - f(z^k)}{f(x^j) - \phi_k(z^k, x^j)} \geqslant \gamma,$$

put $x^{j+1} = z^k$ (serious step), quit inner loop and goto step 8. Otherwise (null step), continue inner loop with step 6.

▷ **Step 6** (`Update working model`). Generate a cutting plane $m_k(\cdot, x^j) = a_k + g_k^\top(\cdot - x^j)$ of f at the null step z^k at counter k belonging to the current serious step x^j. Add (a_k, g_k) to $\mathcal{G}_{k+1}$. Possibly taper out $\mathcal{G}_{k+1}$ by removing some of the older inactive planes in $\mathcal{G}_k$. Build ϕ_{k+1} based on $\mathcal{G}_{k+1}$.

◇ **Step 7** (`Update trust-region radius`). Compute secondary control parameter

$$\widetilde{\rho}_k = \frac{f(x^j) - \phi(z^k, x^j)}{f(x^j) - \phi_k(z^k, x^j)}$$

and put

$$R_{k+1} = \begin{cases} R_k & \text{if } \widetilde{\rho}_k < \widetilde{\gamma}, \\ \frac{1}{2} R_k & \text{if } \widetilde{\rho}_k \geqslant \widetilde{\gamma}. \end{cases}$$

Increase inner loop counter k and loop back to step 4.

◇ **Step 8** (`Update memory radius`). Store new memory radius

$$R_{j+1}^\sharp = \begin{cases} R_k & \text{if } \rho_k < \Gamma, \\ 2R_k & \text{if } \rho_k \geqslant \Gamma. \end{cases}$$

Increase outer loop counter j and loop back to step 2.

Lemma 1. *Let z^k be the trial point at inner loop instant k, associated with the solution y^k of the tangent program, and let g_k be the aggregate subgradient at y^k. Then there exists $\sigma > 0$ depending only on $\theta \in (0, 1)$, M, and the norm $\|\cdot\|$, such that*

$$(5) \qquad\qquad f(x) - \phi_k(z^k, x) \geq \sigma \|g_k\| \|x - z^k\|.$$

Proof. Let $\|\cdot\|$ be the norm used in the trust-region tangent program, $|\cdot|$ the standard Euclidian norm. Since y^k is an optimal solution of (2), we have $0 = g_k + v_k$, where $g_k \in \partial(\phi_k(\cdot, x) + i_C)(y^k)$ and v_k a normal vector to $B(x, R_k)$ at y^k. By the subgradient inequality,

$$g_k^\top(x - y^k) \le \phi_k(x, x) - \phi_k(y^k, x) = f(x) - \phi_k(y^k, x).$$

Now the angle between the vector $y^k - x$ and the normal v_k to the $\|\cdot\|$-ball $B(x, R_k)$ at $y^k \in \partial B(x, R_k)$ is strictly less than $90°$. More precisely, there exists $\sigma' \in (0, 1)$, depending only on the geometry of the ball $B(0, 1)$, such that $\cos \angle(u_k, v_k) \ge \sigma'$ for all such vectors u_k, v_k. But then $g_k^\top(x - y^k) = v_k^\top(y^k - x) \ge \sigma' |v_k||y^k - x| \ge \sigma'' \|v_k\| \|y^k - x\|$ for some $\sigma'' \in (0, 1)$ still depending only on the geometry of the norm $\|\cdot\|$. Invoking (3) for the trial point z^k, and using $\|x - z^k\| \le M \|x - y^k\|$, we get (5) with $\sigma = \sigma'' \theta M^{-1}$. $\qquad \square$

Lemma 2. *Suppose the inner loop at x with trial point z^k at inner loop counter k and solution y^k of the tangent program (2) turns infinitely, and the trust-region radius R_k stays bounded away from 0. Then x is a critical point of (1).*

Proof. We have $\rho_k < \gamma$ for all k. Since $\liminf_{k \to \infty} R_k > 0$, and since the trust-region radius is only reduced when $\widetilde{\rho}_k \ge \widetilde{\gamma}$, and is never increased during the inner loop, we conclude that there exists k_0 such that $\widetilde{\rho}_k < \widetilde{\gamma}$ for all $k \ge k_0$, and also $R_k = R_{k_0} > 0$ for all $k \ge k_0$.

As $z^k, y^k \in B(x, R_{k_0})$, we can extract an infinite subsequence $k \in \mathcal{K}$ such that $z^k \to z$, $y^k \to y$, $k \in \mathcal{K}$. Since we are drawing cutting planes at z^k, we have $\phi_k(z^k, x) = \phi(z^k, x) = m_k(z^k, x)$, and then $\phi_k(z^k, x) \to \phi(z, x)$. Therefore the numerator and denominator in the quotient $\widetilde{\rho}_k$ both converge to $\phi(x, x) - \phi(z, x)$, $k \in \mathcal{K}$. Since $\widetilde{\rho}_k < \widetilde{\gamma} < 1$ for all k, this could only mean $\phi(x, x) - \phi(z, x) = 0$.

Now by condition (3) we have

$$\phi(x, x) - \phi_k(y^k, x) \le \theta^{-1}\left(\phi(x, x) - \phi_k(z^k, x)\right) \to 0,$$

hence $\limsup_{k \in \mathcal{K}} \phi(x, x) - \phi_k(y^k, x) \le 0$. On the other hand, $\phi_k(y^k, x) \le \phi(x, x)$ since y^k solves the tangent program, hence $\phi_k(y^k, x) \to \phi(x, x)$, too.

By the necessary optimality condition for the tangent program (2) there exist $g_k \in \partial_1 \phi_k(y^k, x)$ and a normal vector v_k to $C \cap B(x, R_{k_0})$ at y^k such that $0 = g_k + v_k$. By boundedness of the y^k and local boundedness of the subdifferential, the sequence g_k is bounded, and hence so is the sequence v_k. Passing to yet another subsequence $k \in \mathcal{K}' \subset \mathcal{K}$, we may assume $g_k \to g$, $v_k \to v$, and by upper semi-continuity of the subdifferential, $g \in \partial_1 \phi(y, x)$, and v is in the normal cone to $C \cap B(x, R_{k_0})$ at y. Since $0 = g + v$, we deduce that y is a critical point of the optimization program $\min\{\phi(y, x) : y \in C \cap B(x, R_{k_0})\}$, and since this is a convex program, y is a minimum. But from the previous argument we have seen that $\phi(y, x) = \phi(x, x)$, and since x is admissible for that program, it is also a minimum. A simple convexity argument now shows that x is a minimum of (2). $\qquad \square$

Lemma 3. *Suppose the inner loop at x with trial point z^k and solution y^k of the tangent program at inner loop counter k turns forever, and $\liminf_{k \to \infty} R_k = 0$. Then x is a critical point of (1).*

Proof. This proof uses (5) obtained in Lemma 1. We are in the case where $\widetilde{\rho}_k \ge \widetilde{\gamma}$ for infinitely many $k \in \mathcal{N}$. Since R_k is never increased in the inner loop, we have $R_k \to 0$. Hence $y^k, z^k \to x$ as $k \to \infty$.

We claim that $\phi_k(z^k, x) \to f(x)$. Indeed, we clearly have $\limsup_{k \to \infty} \phi_k(z^k, x) \le \limsup_{k \to \infty} \phi(z^k, x) = \lim_{k \to \infty} \phi(z^k, x) = f(x)$. On the other hand, the exactness plane $m_0(\cdot, x) = f(x) + g_0^\top(\cdot - x)$ is an affine minorant of $\phi_k(\cdot, x)$ at all times k, hence $f(x) = \lim_{k \to \infty} m_0(y^k, x) \le \liminf_{k \to \infty} \phi_k(y^k, x)$, and the two together show $\phi_k(z^k, x) \to f(x)$.

By condition (5) we have $f(x) - \phi_k(z^k, x) \geq \sigma \|g_k\| \|x - z^k\|$, where $g_k \in \partial \left(\phi_k(\cdot, x) + i_C \right)(y^k)$ is the aggregate subgradient. Now assume that $\|g_k\| \geq \eta > 0$ for all k. Then $f(x) - \phi_k(z^k, x) \geq \sigma \eta \|x - z^k\|$.

Since $z^k \to x$, using axiom (M_2) there exist $\epsilon_k \to 0^+$ such that $f(z^k) - \phi(z^k, x) \leq \epsilon_k \|x - z^k\|$. But then

$$\widetilde{\rho}_k = \rho_k + \frac{f(z^k) - \phi(z^k, x)}{f(x) - \phi_k(z^k, x)} \leq \rho_k + \frac{\epsilon_k \|x - z^k\|}{\sigma \eta \|x - z^k\|} = \rho_k + \epsilon_k / (\sigma \eta).$$

Since $\epsilon_k \to 0$, $\rho_k < \gamma$, we have $\limsup_{k \to \infty} \widetilde{\rho}_k \leq \gamma < \widetilde{\gamma}$, contradicting the fact that $\widetilde{\rho}_k > \widetilde{\gamma}$ for infinitely many k. Hence $\|g_k\| \geq \eta > 0$ was impossible.

Select $k \in \mathcal{K}$ such that $g_k \to 0$. Write $g_k = p_k + q_k$ with $p_k \in \partial_1 \phi_k(y^k, x)$ and $q_k \in N_C(y^k)$. Using the boundedness of the y^k extract another subsequence $k \in \mathcal{K}'$ such that $p_k \to p$, $q_k \to q$. Since $y^k \to x$, we have $q \in N_C(x)$. We argue that $p \in \partial f(x)$. Indeed, for any test vector h the subgradient inequality gives

$$p_k^\top h \leq \phi_k(y^k + h, x) - \phi_k(y^k, x) \leq \phi(y^k + h, x) - \phi_k(y^k, x).$$

Since $\phi_k(y^k, x) \to f(x) = \phi(x, x)$, passing to the limit gives

$$p^\top h \leq \phi(x + h, x) - \phi(x, x),$$

proving $p \in \partial_1 \phi(x, x) \subset \partial f(x)$. This proves that x is a critical point of (1). $\qquad \square$

3.2. Convergence of the outer loop. In this section we prove our main convergence result.

Theorem 1. *Suppose f has a strict first-order model ϕ. Let $x^1 \in C$ be such that $\{x \in C : f(x) \leq f(x^1)\}$ is bounded. Let $x^j \in C$ be the sequence of iterates generated by Algorithm 1. Then every accumulation point x^* of the x^j is a critical point of (1).*

Proof. 1) Without loss we consider the case where the algorithm generates an infinite sequence $x^j \in C$ of serious iterates. Suppose that at outer loop counter j the inner loop finds a successful trial step at inner loop counter k_j, that is, $z^{k_j} = x^{j+1}$, where the corresponding solution of the tangent program is $\tilde{x}^{j+1} = y^{k_j}$. Then $\rho_{k_j} \geq \gamma$, which means

$$(6) \qquad f(x^j) - f(x^{j+1}) \geq \gamma \left(f(x^j) - \phi_{k_j}(x^{j+1}, x^j) \right).$$

Moreover, by condition (3) we have $\|\tilde{x}^{j+1} - x^j\| \leq M \|x^{j+1} - x^j\|$ and

$$(7) \qquad f(x^j) - \phi_{k_j}(x^{j+1}, x^j) \geq \theta \left(f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \right),$$

and combining (6) and (7) gives

$$(8) \qquad f(x^j) - f(x^{j+1}) \geq \gamma \theta \left(f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \right).$$

Since $y^{k_j} = \tilde{x}^{j+1}$ is a solution of the k_jth tangent program (2) of the jth inner loop, there exist $g_j \in \partial \left(\phi_{k_j}(\cdot, x^j) + i_C \right)(\tilde{x}^{j+1})$ and a unit normal vector v_j to the ball $B(x^j, R_{k_j})$ at $\tilde{x}^{j+1}$ such that

$$g_j + \|g_j\| v_j = 0.$$

We shall now analyze two types of infinite subsequences, those where the trust-region constraint is active at $\tilde{x}^{j+1}$, and those where it is inactive.

2) Let us start with the simpler case of an infinite subsequence x^j, $j \in J$, where $\|x^j - \tilde{x}^{j+1}\| < R_{k_j}$, i.e., where the trust-region constraint is inactive. There exist $p_j \in \partial_1 \phi_{k_j}(\tilde{x}^{j+1}, x^j)$ and $q_j \in N_C(\tilde{x}^{j+1})$ such that

$$0 = p_j + q_j.$$

By the subgradient inequality, applied to $p_j \in \partial \phi_{k_j}(\cdot, x^j)(\tilde{x}^{j+1})$, we have

$$-q_j^\top(x^j - \tilde{x}^{j+1}) = p_j^\top(x^j - \tilde{x}^{j+1}) \le \phi_{k_j}(x^j, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j)$$
$$= f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \le \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right),$$

using (3). Since $p_j^\top(x^j - \tilde{x}^{j+1}) = q_j^\top(\tilde{x}^{j+1} - x^j) \ge 0$ by Kolmogoroff's inequality, we deduce summability $\sum_{j \in J} p_j^\top(x^j - \tilde{x}^{j+1}) < \infty$, hence $p_j^\top(x^j - \tilde{x}^{j+1}) \to 0$, $j \in J$, and then also $q_j^\top(x^j - \tilde{x}^{j+1}) \to 0$.

Let h be any test vector, then

$$p_j^\top h \le \phi_{k_j}(\tilde{x}^{j+1} + h, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j)$$
$$\le \phi(\tilde{x}^{j+1} + h, x^j) - f(x^j) + f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j)$$
$$\le \phi(\tilde{x}^{j+1} + h, x^j) - f(x^j) + \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right).$$

Now let h' be another test vector and put $h = x^j - \tilde{x}^{j+1} + h'$. Then on substituting this expression we obtain

$$p_j^\top(x^j - \tilde{x}^{j+1}) + p_j^\top h' \le \phi(x^j + h', x^j) - f(x^j) + \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right).$$

Passing to the limit, we have $p_j^\top(x^j - \tilde{x}^{j+1}) \to 0$ by the above, and $f(x^j) - f(x^{j+1}) \to 0$ by the construction of the descent method. Moreover, $\limsup_{j \in J} \phi(x^j + h', x^j) \le \phi(x^* + h', x^*)$ by axiom (M_3) and $p_j \to p$ for some p. That shows

$$p^\top h' \le \phi(x^* + h', x^*) - f(x^*) = \phi(x^* + h', x^*) - \phi(x^*, x^*).$$

Since h' was arbitrary and $\phi(\cdot, x^*)$ is convex, we deduce $p \in \partial_1 \phi(x^*, x^*)$, hence $p \in \partial f(x^*)$ by axiom (M_1).

Now observe that $\tilde{x}^{j+1} \to \tilde{x}$ and $q_j \to q \in N_C(\tilde{x})$. We wish to show that $q \in N_C(x^*)$. Since $q_j^\top(x^j - \tilde{x}^{j+1}) \to 0$, we have $q^\top(x^* - \tilde{x}) = 0$, but $q \neq 0$ and $x^* - \tilde{x} \neq 0$. Now for any element $x \in C$ we have $q^\top(\tilde{x} - x) \ge 0$ by Kolmogoroff's inequality. Hence $q^\top(x^* - x) = q^\top(\tilde{x} - x) + q^\top(x^* - \tilde{x}) = q^\top(\tilde{x} - x) \ge 0$, so Kolmogoroff's inequality holds also at x^*, proving $q \in N_C(x^*)$. We have shown that $0 = p + q \in \partial\left(\phi(\cdot, x^*) + i_C\right)(x^*)$, hence x^* is a critical point of (1).

3) Let us now consider the more complicated case of an infinite subsequence, where $\|x^j - \tilde{x}^{j+1}\| = R_{k_j}$ with $g_j \neq 0$. In other words, the trust-region constraint is active at $\tilde{x}^{j+1}$. Passing to a subsequence, we may assume $x^j \to x^*$, and we have to show that x^* is critical.

Let u_j be the unit vector $u_j = (\tilde{x}^{j+1} - x^j)/\|\tilde{x}^{j+1} - x^j\|$. Then if the norm $\|\cdot\|$ coincides with the Euclidian norm $|\cdot|$, we have $u_j = v_j$. For other norms this is no longer the case, but for any such norm there exists $\sigma > 0$ such that $u_j^\top v_j \ge \sigma > 0$ for all j. Then

$$g_j^\top(x^j - \tilde{x}^{j+1}) = -\|x^j - \tilde{x}^{j+1}\|g_j^\top u_j = \|x^j - \tilde{x}^{j+1}\|\|g_j\|v_j^\top u_j \ge \sigma\|g_j\|\|x^j - \tilde{x}^{j+1}\|.$$

By the subgradient inequality, and using $x^j, \tilde{x}^{j+1} \in C$, we have

$$g_j^\top(x^j - \tilde{x}^{j+1}) \le \phi_{k_j}(x^j, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) = f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j).$$

Altogether

$$(9) \qquad f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \ge \sigma\|g_j\|\|x^j - \tilde{x}^{j+1}\|.$$

Combining this with (8) gives

$$\|g_j\|\|x^j - \tilde{x}^{j+1}\| \le \sigma^{-1}\gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right).$$

Summing both sides from $j = 1$ to $j = J$ gives

$$\sum_{j=1}^{J} \|g_j\| \|x^j - \tilde{x}^{j+1}\| \leq \sigma^{-1} \gamma^{-1} \theta^{-1} \left(f(x^1) - f(x^{J+1}) \right).$$

Since the values $f(x^j)$ are decreasing and $\{x \in C : f(x) \leq f(x^1)\}$ is bounded, the sequence x^j must be bounded. We deduce that the right hand side is bounded, hence the series on the left converges:

$$(10) \qquad\qquad \sum_{j=1}^{\infty} \|g_j\| \|x^j - \tilde{x}^{j+1}\| < \infty.$$

In particular, this implies $\|g_j\| \|x^j - \tilde{x}^{j+1}\| \to 0$. Using $\|x^j - x^{j+1}\| \leq M \|x^j - \tilde{x}^{j+1}\|$, we also have $\|g_j\| \|x^j - x^{j+1}\| \to 0$.

We shall now have to distinguish two subcases. Either there exists a subsequence $J' \subset J$ such that $R_{k_j} \to 0$ as $j \in J'$, or $R_{k_j} \geq R_0 > 0$ for all $j \in J$. The second subcase is discussed in 4) below, the first is handled in 5) - 6).

4) Let us consider the sub-case of an infinite subsequence $j \in J$ where $\|x^j - \tilde{x}^{j+1}\| = R_{k_j} \geq R_0 > 0$ for every $j \in J$. Going back to (10), we see that we now must have $g_j \to 0$, as $x^j - \tilde{x}^{j+1} \not\to 0$. Let us write $g_j = p_j + q_j$, where $p_j \in \partial_1 \phi_{k_j}(\tilde{x}^{j+1}, x^j)$ and $q_j \in N_C(\tilde{x}^{j+1})$. Then

$$p_j^\top (x^j - \tilde{x}^{j+1}) \leq \phi_{k_j}(x^j, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \leq \gamma^{-1} \theta^{-1} \left(f(x^j) - f(x^{j+1}) \right).$$

Now $g_j^\top (x^j - \tilde{x}^{j+1}) = p_j^\top (x^j - \tilde{x}^{j+1}) + q_j^\top (x^j - \tilde{x}^{j+1}) \leq p_j^\top (x^j - \tilde{x}^{j+1})$, because Kolmogoroff's inequality for $\tilde{x}^{j+1} \in C$ and $q_j \in N_C(\tilde{x}^{j+1})$ gives $q_j^\top (\tilde{x}^{j+1} - x^j) \geq 0$. Hence we have

$$g_j^\top (x^j - \tilde{x}^{j+1}) \leq p_j^\top (x^j - \tilde{x}^{j+1}) \leq \gamma^{-1} \theta^{-1} \left(f(x^j) - f(x^{j+1}) \right),$$

so $p_j^\top (x^j - \tilde{x}^{j+1}) \to 0$, because the lefthand term and the righthand term both converge to 0. As a consequence, we also have $q_j^\top (x^j - \tilde{x}^{j+1}) \to 0$.

Now observe that the sequence $x^j \in C$ is also bounded, because $\{x \in C : f(x) \leq f(x^1)\}$ is bounded and the x^j form a descent sequence for f. Let us say $\|x^1 - x^j\| \leq K$ for all j. We argue that the p_j are then also bounded. This can be shown as follows. Let h be a test vector with $\|h\| = 1$. Then

$$\begin{aligned}
p_j^\top h &\leq \phi_{k_j}(\tilde{x}^{j+1} + h, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\
&\leq \phi(\tilde{x}^{j+1} + h, x^j) - m_{0j}(\tilde{x}^{j+1}, x^j) \\
&= \phi(\tilde{x}^{j+1} + h, x^j) - f(x^j) - g_{0j}^\top (\tilde{x}^{j+1} - x^j) \\
&\leq C + |f(x^1)| + \|g_{0j}\| \|x^j - \tilde{x}^{j+1}\|,
\end{aligned}$$

where $C := \max\{\phi(u, v) : \|u - x^1\| \leq MK + 1, \|v - x^1\| \leq K\} < \infty$ and where $g_{0j} \in \partial f(x^j)$ by the definition of the exactness plane at x^j. But observe that ∂f is locally bounded by [37], so $\|g_{0j}\| \leq K' < \infty$. We deduce $\|p_j\| \leq C + |f(x^1)| + K'(2K + M) < \infty$. Hence the sequence p_j is bounded, and since $g_j = p_j + q_j \to 0$ by the above, the sequence q_j is also bounded.

Therefore, on passing to a subsequence $j \in J'$, we may assume $x^j \to x^*$, $\tilde{x}^{j+1} \to \tilde{x}$, $p_j \to p$, $q_j \to q$. Then $q \in N_C(\tilde{x})$. Now from the subgradient inequality

$$\begin{aligned}
p_j^\top h &\leq \phi_{k_j}(\tilde{x}^{j+1} + h, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\
&\leq \phi(\tilde{x}^{j+1} + h, x^j) - f(x^j) + f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\
&\leq \phi(\tilde{x}^{j+1} + h, x^j) - \phi(x^j, x^j) + \gamma^{-1} \theta^{-1} \left(f(x^j) - f(x^{j+1}) \right),
\end{aligned}$$

where we use (3), $\phi_{k_j} \leq \phi$, and acceptance $\rho_{k_j} \geq \gamma$, and where the test vector h is arbitrary. Let h' another test vector and put $h = x^j - \tilde{x}^{j+1} + h'$. Substituting this gives

$$(11) \qquad p_j^\top (x^j - \tilde{x}^{j+1}) + p_j^\top h' \leq \phi(x^j + h', x^j) - \phi(x^j, x^j) + \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right).$$

Now $p_j^\top(x^j - \tilde{x}^{j+1}) = (p_j + q_j)^\top(x^j - \tilde{x}^{j+1}) + q_j^\top(\tilde{x}^{j+1} - x^j) \geq (p_j + q_j)^\top(x^j - \tilde{x}^{j+1})$ using Kolmogoroff's condition for $q_j \in N_C(\tilde{x}^{j+1})$. Therefore, on passing to the limit in (11), using $(p_j+q_j)^\top(x^j-\tilde{x}^{j+1}) \to 0$, $f(x^j)-f(x^{j+1}) \to 0$, $p_j \to p$ and $\limsup_{j \in J'} \phi(x^j+h',x^j) \leq \phi(x^*+h',x^*)$, which follows from axiom (M_3), we find

$$p^\top h' \leq \phi(x^* + h', x^*) - \phi(x^*, x^*).$$

Since h' was arbitrary, we deduce $p \in \partial_1 \phi(x^*, x^*)$, and by axiom (M_1), $p \in \partial f(x^*)$.

It remains to show $q \in N_C(x^*)$. Now recall that $q_j^\top(x^j - \tilde{x}^{j+1}) \to 0$ was shown at the beginning of part 4), so $q^\top(x^* - \tilde{x}) = 0$. Given any test element $x \in C$, Kolmogoroff's inequality for $q \in N_C(\tilde{x})$ gives $q^\top(\tilde{x}-x) \geq 0$. But then $q^\top(x^*-x) = q^\top(\tilde{x}-x)+q^\top(x^*-\tilde{x}) = q^\top(\tilde{x}-x) \geq 0$, so Kolmogoroff's inequality also holds for q at x^*, proving $q \in N_C(x^*)$.

With $q \in N_C(x^*)$ and $g = p + q = 0$, we have shown that x^* is a critical point of (1). That settles the case where the trust-region radius is active and bounded away from 0.

5) It remains to discuss the most complicated sub-case of an infinite subsequence $j \in J$, where the trust-region constraint is active and $R_{k_j} \to 0$. This needs two sub-sub-cases. The first of these is a sequence $j \in J$ where in each jth outer loop the trust-region radius was reduced at least once. The second sub-sub-case are infinite subsequences where the trust-region radius stayed frozen $(R_j^\sharp = R_{k_j})$ throughout the jth inner loop for every $j \in J$. This is discussed in 6) below.

Let us first consider the case of an infinite sequence $j \in J$ where R_{k_j} is active at $\tilde{x}^{j+1}$, and $R_{k_j} \to 0$, $j \in J$, such that during the jth inner loop the trust-region radius was reduced at least once. Suppose this happened the last time before acceptance at inner loop counter $k_j - \nu_j$. Then for $j \in J$,

$$R_{k_j} = R_{k_j-1} = \cdots = R_{k_j-\nu_j} = \tfrac{1}{2} R_{k_j-\nu_j-1}.$$

By step 7 of the algorithm, that implies

$$\tilde{\rho}_{k_j-\nu_j} \geq \tilde{\gamma}, \quad \rho_{k_j-\nu_j} < \gamma.$$

Now $\|x^{j+1} - x^j\| \leq R_{k_j}$ and $\|z^{k_j-\nu_j} - x^j\| \leq R_{k_j-\nu_j-1} = 2R_{k_j}$, hence $x^{j+1} - z^{k_j-\nu_j} \to 0$, $x^j - z^{k_j-\nu_j} \to 0$, $j \in J''$. From axiom $(\widetilde{M_2})$ we deduce that there exists a sequence $\epsilon_j \to 0^+$ such that

$$f(z^{k_j-\nu_j}) \leq \phi(z^{k_j-\nu_j}, x^j) + \epsilon_j \|z^{k_j-\nu_j} - x^j\|.$$

By the definition of the aggregate subgradient $\tilde{g}_j \in \partial\left(\phi_{k_j-\nu_j}(\cdot, x^j) + i_C\right)(y^{k_j-\nu_j})$ and Lemma 1 we have $f(x^j) - \phi_{k_j-\nu_j}(z^{k_j-\nu_j}, x^j) \geq \sigma \|\tilde{g}_j\|\|x^j - z^{k_j-\nu_j}\|$.

Recall that $x^j \to x^*$ and that we have to show that x^* is critical. It suffices to show that there is a subsequence $j \in J'$ with $g_j \to 0$. Assume on the contrary that $\|\tilde{g}_j\| \geq \eta > 0$ for every $j \in J$. Then

$$f(x^j) - \phi_{k_j-\nu_j}(z^{k_j-\nu_j}, x^j) \geq \eta\sigma \|z^{k_j-\nu_j} - x^j\|.$$

Now

$$\tilde{\rho}_{k_j-\nu_j} = \rho_{k_j-\nu_j} + \frac{f(z^{k_j-\nu_j}) - \phi(z^{k_j-\nu_j}, x^j)}{f(x^j) - \phi_{k_j-\nu_j}(z^{k_j-\nu_j}, x^j)} \leq \rho_{k_j-\nu_j} + \frac{\epsilon_j\|z^{k_j-\nu_j} - x^j\|}{\eta\|z^{k_j-\nu_j} - x^j\|} < \tilde{\gamma}$$

for $j \in J$ sufficiently large, contradicting $\tilde{\rho}_{k_j-\nu_j} \geq \tilde{\gamma}$. This shows that there must exist a subsequence J' such that $\tilde{g}_j \to 0$, $j \in J'$. Passing to the limit $j \in J'$, this shows $0 \in \partial\left(\phi(\cdot, x^*) + i_C\right)(x^*)$, hence x^* is critical for (1).

6) Now consider an infinite subsequence $j \in J$ where $x^j \to x^*$, the trust-region radius R_{k_j} was active at $\tilde{x}^{j+1}$ when x^{j+1} was accepted, $R_{k_j} \to 0$, but during the jth inner loop the trust-region radius was never reduced. In the classical case this can only happen when x^{j+1} at j is immediately accepted, but with bundling this could also happen when the inner loop adds cutting planes for a time, while the test in step 7 keeps $R_{k+1} = R_k$ in the inner loop. Since $R_{k_j} \to 0$, the work to bring the radius to 0 must be put about somewhere else. For every $j \in J$ define $j' \in \mathbb{N}$ to be the largest index $j' < j$ such that in the j'th inner loop, the trust-region radius was reduced at least once. Let $J' = \{j' : j \in J\}$, where we understand $j \mapsto j'$ as a function. Passing to a subsequence of J, J', we may assume that $x^{j'} \to x'$ and $g_{j'} \to 0$, because the sequence J' corresponds to one of the cases discussed in parts 2) - 5). Passing to jet another subsequence, we may arrange that the sequences J, J' are interlaced. That is, $j' < j < j'^+ < j^+ < j'^{++} < j^{++} < \cdots \to \infty$. This is because j' tends to ∞ as a function of j.

Now assume that there exists $\eta > 0$ such that $\|g_j\| \geq \eta$ for all $j \in J$. Then since $x^j \to x^*$, we also have $x^{j+1} \to x^*$. Fix $\epsilon > 0$ with $\epsilon < \eta$. For $j \in J$ large enough we have $\|g_{j'}\| < \epsilon$, because $g_{j'} \to 0$, $j' \in J'$, and as j gets larger, so does j'. That means in the interval $[j', j)$ there exists an index $j'' \in \mathbb{N}$ such that

$$\|g_{j''}\| < \epsilon, \quad \|g_i\| \geq \epsilon \text{ for all } i = j'' + 1, \ldots, j.$$

The index j'' may coincide with j', it might also be larger, but it precedes j. In any case, $j \mapsto j''$ is again a function on J and defines another infinite index set J'' still interlaced with J.

Now recall from part 3), estimate (10), and $\|x^j - x^{j+1}\| \leq M\|x^j - \tilde{x}^{j+1}\|$, that for some constant $c > 0$

$$\sum_{i=j''+1}^{j} \|g_i\|\|x^i - x^{i+1}\| \leq c \left(f(x^{j''+1}) - f(x^{j+1}) \right) \to 0 \qquad (j \in J, j \to \infty, j \mapsto j'').$$

Since by construction $\|g_i\| \geq \epsilon$ for all $i \in [j'' + 1, \ldots, j]$, and that for all $j \in J$, the sequence $\sum_{i=j''+1}^{j} \|x^i - x^{i+1}\| \to 0$ converges as $j \in J, j \to \infty$, and by the triangle inequality, $x^{j''+1} - x^{j+1} \to 0$. Therefore $x^{j''+1} \to x^*$. Since $g_{j''} \in \partial(f + i_C)(x^{j''+1})$, passing to yet another subsequence and using upper semi-continuity of the subdifferential, we get $g_{j''} \to \tilde{g} \in \partial(f + i_C)(x^*)$. Since $\|g_{j''}\| < \epsilon$, we have $\|\tilde{g}\| \leq \epsilon$. It follows that $\partial(f + i_C)(x^*)$ contains an element of norm $\leq \epsilon$. As $\epsilon < \eta$ was arbitrary, we conclude that $0 \in \partial(f + i_C)(x^*)$. That settles the remaining case. $\square$

4. Stopping test

A closer look at the convergence proof indicates stopping criteria for algorithm 1. As is standard in bundle methods, step 2 is not executed as such but delegated to the inner loop. When a serious step x^{j+1} is accepted, we apply the tests

$$\frac{\|x^j - x^{j+1}\|}{1 + \|x^j\|} < \text{tol}_1, \quad \frac{f(x^j) - f(x^{j+1})}{1 + |f(x^j)|} < \text{tol}_2$$

in tandem with

$$\frac{\min\{\|g_j\|, \|g_{j'}\|\}}{1 + |f(x^j)|} < \text{tol}_3.$$

Here g_j is the aggregate subgradient at acceptance. In the case treated in part 6) of the proof we had to consider the largest index $j' < j$, where the trust-region radius was reduced for the last time. If in the inner loop at x^j leading to x^{j+1} the trust-region radius was not reduced, we have to consider both aggregates, otherwise $\|g_j\|/(1 + \|x^j\|) < \text{tol}_3$ suffices. If the three criteria are satisfied, then we return x^{j+1} as optimal.

On the other hand, when the inner loop has difficulties finding a new serious iterate, and if a maximum number $k_{\max}$ is exceeded, or if for $\nu_{\max}$ consecutive steps

$$\frac{\|x^j - z^k\|}{1 + \|x^j\|} < \text{tol}_1, \quad \frac{f(x^j) - f(z^k)}{1 + |f(x^j)|} < \text{tol}_2$$

in tandem with

$$\frac{\|g_k\|}{1 + |f(x^j)|} < \text{tol}_3$$

are satisfied, where g_k is the aggregate subgradient at y^k, then the inner loop is stopped and x^j is returned as optimal. In our tests we use $k_{\max} = 50$, $\nu_{\max} = 5$, $\text{tol}_1 = \text{tol}_2 = 10^{-5}$, $\text{tol}_3 = 10^{-6}$. Typical values in algorithm 1 are $\gamma = 0.0001$, $\widetilde{\gamma} = 0.0002$, $\Gamma = 0.1$.

5. Applications

In this section we highlight the potential of the model-based trust-region approach by presenting several applications.

5.1. Full model versus working model. Our convergence theory covers the specific case $\phi_k = \phi$, which we call the *full model case*. Here the algorithm simplifies, because cutting planes are redundant, so that step 6 becomes obsolete. Moreover, in step 7 the quotient $\widetilde{\rho}_k$ always equals 1, so the only action taken is reduction of the trust-region radius. This is now close to the rationale of the classical trust-region method.

5.2. Natural model. For a composite function $f = g \circ F$ with g convex and F of class C^1 the *natural model* is $\phi(y, x) = g\left(F(x) + F'(x)(y - x)\right)$, because it is strict and can be used in algorithm 1. In the full model case $\phi_k = \phi$, our algorithm reduces to the algorithm of Ruszczyński [38, Chap. 7.5] for composite nonsmooth functions.

5.3. Spectral model. An important field of applications, where the natural model often comes into action, are eigenvalue optimization problems

$$\begin{array}{ll} (12) & \begin{array}{ll} \text{minimize} & \lambda_1\left(\mathcal{F}(x)\right) \\ \text{subject to} & x \in C \end{array} \end{array}$$

where $\mathcal{F} : \mathbb{R}^n \to \mathbb{S}^m$ is a class C^1-mapping into the space of $m \times m$ symmetric or Hermitian matrices $\mathbb{S}^m$, and $\lambda_1(\cdot)$ the maximum eigenvalue function on $\mathbb{S}^m$, which is convex but nonsmooth. Here the natural model is $\phi(y, x) = \lambda_1\left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right)$, where $\mathcal{F}'$ is the differential of $\mathcal{F}$. Note that nonlinear semidefinite programs

$$\begin{array}{ll} (13) & \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \mathcal{F}(x) \preceq 0 \\ & x \in C \end{array} \end{array}$$

are special cases of (12) if we use exact penalization and write (13) in the form

$$\begin{array}{ll} \text{minimize} & f(x) + c \max\{0, \lambda_1\left(\mathcal{F}(x)\right)\} \\ \text{subject to} & x \in C \end{array}$$

with a suitable $c > 0$. Namely, this new objective may be written as the maximum eigenvalue of the mapping

$$\mathcal{F}^\sharp(x) = \left[\begin{array}{cc} f(x) & 0 \\ 0 & f(x)I_m + c\mathcal{F}(x) \end{array} \right] \in \mathbb{S}^{1+m}.$$

Let us apply the bundling idea to (12) using the natural model ϕ. Here we may build working models ϕ_k generated by infinite sets $\mathcal{G}_k$ of cuts (a, g) from ϕ, and still arrive at a computable tangent program. Indeed, suppose y^k is a null step at serious iterate x. According to step 6 of algorithm 1 we have to generate one or several cutting planes at y^k.

This means we have to compute $g_k \in \partial\lambda_1 \left(\mathcal{F}(x) + \mathcal{F}'(x)(\cdot - x)\right)(y^k)$. Now by the generalized chain rule the subdifferential of the composite function $y \mapsto \lambda_1 \left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right)$ at y is $\mathcal{F}'(x)^* \partial\lambda_1 \left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right)$, where $\partial\lambda_1$ is now the convex subdifferential of λ_1 in matrix space $\mathbb{S}^m$, i.e.,

$$\partial\lambda_1(X) = \{G \in \mathbb{S}^m : G \succeq 0, \operatorname{tr}(G) = 1, G \bullet X = \lambda_1(X)\}$$

with $X \bullet Y = \operatorname{tr}(XY)$ the scalar product in $\mathbb{S}^m$. Here $\mathcal{F}'(x)^* : \mathbb{S}^m \to \mathbb{R}^n$ is the adjoint of the linear operator $\mathcal{F}'(x)$. It follows that every subgradient g of the composite function is of the form

$$(14) \qquad g = \mathcal{F}'(x)^* G, \quad G \in \partial\lambda_1 \left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right).$$

The corresponding a is $a = \lambda_1 \left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right) + g^\top(x - y)$. As soon as the maximum eigenvalue $\lambda_1(X)$ has multiplicity > 1, the set $\partial\lambda_1(X)$ is not singleton, and we may therefore add the entire subdifferential to the new set $\mathcal{G}_{k+1}$.

Let y^k be a null step, and let Q_r be an $m \times t_k$ matrix whose t_k columns form an orthogonal basis of the maximum eigenspace of $\mathcal{F}(x) + \mathcal{F}'(x)(y^k - x)$. Let Y_k be a $t_k \times t_k$-matrix with $Y_k = Y_k^\top$, $Y_k \succeq 0$, $\operatorname{tr}(Y_k) = 1$, then subgradients (14) are of the form $G_k = Q_k Y_k Q_k^\top$. Therefore all pairs $(a_r, g_r(Y_r)) \in \mathcal{G}_k$ are of the form

$$a_r = \lambda_1 \left(\mathcal{F}(x) + \mathcal{F}'(x)(y^r - x)\right), \quad g_r(Y_r) = \mathcal{F}'(x)^* G_r, \quad G_r = Q_r Y_r Q_r^\top,$$

indexed by $Y_r \succeq 0$, $\operatorname{tr}(Y_r) = 1$, $Y_r \in \mathbb{S}^{t_r}$ stemming from older null steps $r = 1, \ldots, k$. The trust-region tangent program is then

$$\begin{aligned} \text{minimize} \quad & \max_{r=1,\ldots,k} a_r + \lambda_1 \left(Q_r \mathcal{F}'(x)(y - y^r)Q_r^\top\right) \\ \text{subject to} \quad & y \in C, \ \|y - x\| \le R \end{aligned}$$

This is a linear semidefinite program if a polyhedral or a conical norm is used, and if C is a convex semidefinite constraint set.

We can go one step further and consider semi-infinite maximum eigenvalue problems as in [7], as this has scope for applications in automatic control. It allows us for instance to optimize the H_∞-norm, or more general IQC-constrained programs, see [6].

5.4. Standard model. The most straightforward choice of a model is the *standard model*

$$\phi^\sharp(y, x) = f(x) + f^\circ(x, y - x),$$

as it gives a direct substitute for the first-order Taylor expansion of f at x. Here the full model tangent program (2) has the specific form

$$(15) \qquad \begin{aligned} \text{minimize} \quad & f(x) + f^\circ(x, y - x) \\ \text{subject to} \quad & y \in C \\ & \|y - x\| \le R_k \end{aligned}$$

and if a polyhedral working model $\phi_k^\sharp$ is used to approximate $\phi^\sharp$ via bundling, then we get an even simpler tangent program of the form

$$(16) \qquad \begin{aligned} \text{minimize} \quad & f(x) + \max_{i=1,\ldots,k} g_i^\top(y - x) \\ \text{subject to} \quad & y \in C \\ & \|y - x\| \le R_k \end{aligned}$$

where $g_i \in \partial f(x)$. If a polyhedral norm is used and C is a polyhedron, then (16) is just a linear program, which makes this line attractive computationally.

Remark 11. Consider the unconstrained case $C = \mathbb{R}^n$ with $\phi_k^\sharp = \phi^\sharp$, then $y^k = x - R_k g(x)/\|g(x)\|$, where $g(x) = \underset{g \in \partial f(x)}{\operatorname{argmin}} \{\|g\| : g \in \partial f(x)\}$, and this is the nonsmooth steepest descent step of length R_k at x. In classical trust-region algorithms the steepest descent step of length R_k is often chosen as the first-order Cauchy step.

This raises the following natural question. Can we use the solution of y^k of (15), or (16), as a nonsmooth Cauchy point? Since we do not want to keep the reader on the tenterhooks too long, here is the answer: *no we can't*. Namely, in order to be allowed to use the standard model in Algorithm 1, and the solution of (15), (16) as a Cauchy point for other models, $\phi^\sharp$ has to be strict, because this is required in Theorem 1. A sufficient condition for strictness of $\phi^\sharp$ is given in [32]. We need the following

Definition 4 (Spingarn [44], Rockafellar-Wets [37]). A locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ is lower-C^1 at $x_0 \in \mathbb{R}^n$ if there exist a compact space $\mathbb{K}$, a neighborhood U of x_0, and a mapping $F : \mathbb{R}^n \times \mathbb{K} \to \mathbb{R}$ such that

$$(17) \qquad\qquad f(x) = \max_{y \in \mathbb{K}} F(x, y)$$

for all $x \in U$, and F and $\partial F/\partial x$ are jointly continuous. The function f is said to be upper-C^1 at x_0 if $-f$ is lower-C^1 at x_0. $\qquad\square$

Lemma 4. *(See [32]). Suppose f is locally Lipschitz and upper C^1. Then the standard model $\phi^\sharp$ of f is strict.* $\qquad\square$

Example 1. The lightning function $f : \mathbb{R} \to \mathbb{R}$ in [25] is an example where $\phi^\sharp$ is strict, but f is not upper C^1. It is Lipschitz with constant 1 and has $\partial f(x) = [-1, 1]$ for every x. The standard model of f is strict, because for all x, y there exists $\rho = \rho(x, y) \in [-1, 1]$ such that

$$f(y) = f(x) + \rho|y - x| \leq f(x) + \operatorname{sign}(y - x)(y - x)$$
$$\leq f(x) + f^\circ(x, y - x) = \phi^\sharp(x, y - x),$$

using the fact that $\operatorname{sign}(y - x) \in \partial f(x)$. At the same time f is certainly not upper-C^1, because it is not semi-smooth in the sense of [28].

When using the standard model $\phi^\sharp$ in Algorithm 1, we expect the trust-region method to coincide with its classical antecedent, or at least, to be very similar to it. But we expect more! Let $\mathscr{S}$ be the class of nonsmooth locally Lipschitz functions f which have a strict standard model $\phi^\sharp$. Suppose a subclass $\mathscr{S}'$ of $\mathscr{S}$ leads to simplifications of algorithm 1 which reduce it to its classical alter ego. Then we have a theoretical justification to say that functions $f \in \mathscr{S}'$, even though nonsmooth, can be optimized *as if they were smooth*.

Following Borwein and Moors [9], a function f is called *essentially smooth* if it is locally Lipschitz and strictly differentiable almost everywhere. The lightning function of example 1 is a pathological case, which is differentiable almost everywhere, but nowhere strictly differentiable. In practice we expect nonsmooth functions to be essentially smooth. This is for instance the case for semi-smooth functions in the sense of [28], for arc-wise essentially smooth functions, or for pseudo-regular functions in the sense of [9].

Proposition 2. *Let f be essentially smooth. Let $x^1 \in C$ be such that $\{x \in C : f(x) \leq f(x^1)\}$ is bounded. Suppose the standard model $\phi^\sharp$ is used in algorithm 1. Let trial points $z^k \in C$ satisfying (3) in step 4 are drawn at random and independently according to a continuous probability distribution on C. Then with probability one the steps of the algorithm are identical with the steps of the classical trust-region algorithm. Moreover, if $\phi^\sharp$ is strict, then every accumulation point of the sequence x^j is critical.*

Proof. Since there exists a full neighborhood U of y^k such that every $z^k \in U \cap C$ is a valid trial point, and since the elements in $U \cap C$ are with probability 1 points of strict differentiability, the entire sequence x^j consists with probability 1 of points of strict differentiability. $\qquad \square$

Note that we should not expect the y^k themselves to be points of differentiability, let alone strict differentiability. In fact the y^k will typically lie in a set of measure 0. For instance, if C is a polyhedron, then y^k is typically a vertex of C, or a vertex of the polyhedron of the linear program (16).

Proposition 2 applies in particular when f is upper C^1, because upper C^1-functions are essentially smooth. However, for upper C^1 functions we have the following stronger result. A similar observation in the context of bundle methods was first made in [15].

Lemma 5. *Suppose f is locally Lipschitz and upper-C^1 and the standard model $\phi^\sharp$ is used in algorithm 1. Then we can choose the cutting plane $m_k(\cdot, x) = f(x) + g_k^\top(\cdot - x)$ in step 6 with $g_k \in \partial f(x)$ arbitrarily, because $f^\circ(x, z^k - x) - g_k^\top(z^k - x) \le \epsilon_k \|z^k - x\|$ holds automatically for some $\epsilon_k \to 0^+$ in the inner loop at x, and $f^\circ(x^j, x^{j+1} - x^j) - g_j^\top(x^{j+1} - x^j) \le \epsilon_j \|x^{j+1} - x^j\|$ holds automatically for some $\epsilon_j \to 0^+$ in the outer loop.*

Proof. Daniilidis and Georgiev [14, Thm. 2] prove that an upper C^1 function is super-monotone at x in the following sense: For every $\epsilon > 0$ there exists $\delta > 0$ such that $(g_1 - g_2)^\top(x_1 - x_2) \le \epsilon \|x_1 - x_2\|$ for all $x_i \in U$ and $g_i \in \partial f(x_i)$. Hence for sequences $x^j, y^j \to x$ we find $\epsilon_j \to 0^+$ such that $(g_j^* - g_j)^\top(x^j - y^k) \le \epsilon_j \|y^j - x^j\|$ for all $g_j^* \in \partial f(y^j)$, $g_j \in \partial f(x^j)$. Choosing g_j^* such that $f^\circ(x^j, y^j - x^j) = g_j^{*\top}(y^j - x^j)$ then gives the result. $\qquad \square$

As a consequence we have the following

Theorem 2. *Suppose f is upper-C^1, $x^1 \in C$, and $\{x \in C : f(x) \le f(x^1)\}$ is bounded. Suppose the classical trust-region algorithm is used, that is, the only cutting plane in step 6 chosen at x is an arbitrarily exactness plane, and in step 7 the trust-region radius is reduced whenever a null step occurs. Then every accumulation point of the sequence of serious iterates x^j is a critical point of (1). Moreover, if f satisfies the Kurdyka-Łojasiewicz inequality, then the x^j converge to a single critical point x^* of f.*

Proof. By Lemma 5 the proof of Theorem 1 applies regardless how we choose cutting planes from $\phi^\sharp$. We exploit this by choosing them in the simplest possible way, namely we take only one exactness plane and keep it all the time. If f is differentiable at x then our only choice is $m(\cdot, x) = f(x) + \nabla f(x)^\top(\cdot - x)$, otherwise we take $m(\cdot, x) = f(x) + g^\top(\cdot - x)$ with an arbitrary $g \in \partial f(x)$. This makes step 6 redundant and reduces step 7 to the usual modification of the trust-region radius. And this is now just the classical trust-region strategy, for which we then have subsequence convergence by Theorem 1.

It remains to show that under the Kurdyka-Łojasiewicz inequality the x^j converge even to a single limit. This can be based on the technique of [1, 8, 32]. $\qquad \square$

Remark 12. An axiomatic approach to trust-region methods is Dennis *et al.* [17], and the idea is adopted in [13, Chap. 11]. The difference with our approach is that ϕ in [17, 13] has to be jointly continuous, while we use the weaker axiom (M_3), and that their f has to be regular, which precludes the use of the standard model $\phi^\sharp$, hence makes it impossible to use the Cauchy point. Bundling is not discussed in these approaches.

On the other hand, the authors of [17], [13] do allow non-convex models, while in our approach $\phi(\cdot, x)$ is convex because we want to assure a computable tangent program, and be able to draw cutting planes. Convexity of $\phi(\cdot, x)$ could be relaxed to $\phi(\cdot, x)$ being lower C^1. For that the downshift idea [28, 31] would have to be used.

5.5. Delamination problem. Contact mechanics is a domain where nonsmooth optimization programs arise frequently. When potential energy is minimized under non-monotone friction laws, then programs with lower-C^1 functions arise. On the other hand, quasi-static delamination problems lead to minimization of upper-C^1 criteria, see [16, 36, 2] for more information.

5.6. Model for splitting. Suppose we wish to optimize a function $f = g + h$ where g is differentiable and h is convex. Then a model ϕ for f is $\phi(y, x) = g(x) + \nabla g(x)^\top (y - x) + h(y) = \phi_g^\sharp(y, x) + h(y)$. Indeed, for the differentiable g the first-order Taylor expansion is natural, and the convex h is its own strict model. Cutting planes are now sums of cutting planes of the two model components. Algorithm 1 based on ϕ could then be an alternative to a splitting technique, in particular, as ours carries over easily to the case when h is lower-C^2.

5.7. Failure of the Cauchy point. We will show by way of an example that the classical trust-region approach based on the Cauchy point fails in the nonsmooth case. We operate algorithm 1 with the full standard model $\phi^\sharp$, compute the Cauchy point y^k via (15) based on the Euclidian norm, and use $z^k = y^k$ as the trial step. This corresponds essentially to a classical first-order trust-region method.

The following example adapted from [23] can be used to show the difficulties with this classical scheme. We define a convex piecewise affine function $f : \mathbb{R}^2 \to \mathbb{R}$ as

$$f(x) = \max\{f_0(x), f_{\pm 1}(x), f_{\pm 2}(x)\}$$

where $x = (x_1, x_2)$ and

$$f_0(x) = -100, f_{\pm 1}(x) = \pm 2x_1 + 3x_2, f_{\pm 2}(x) = \pm 5x_1 + 2x_2.$$

The plot below shows that part of the level curve $[f = a]$ which lies in the upper half plane $x_2 \geq 0$. It consists of the polygon connecting the five points $(-\frac{a}{5}, 0)$, $(-\frac{a}{11}, \frac{3a}{11})$, $(0, \frac{a}{3})$, $(\frac{a}{11}, \frac{3a}{11})$, $(\frac{a}{5}, 0)$. We are interested in that part of the lower level set $[f \leq a]$, which lies within the gray-shaded dragon-shaped area inside the polygon $[f \leq a]$, and above the x_1-axis.
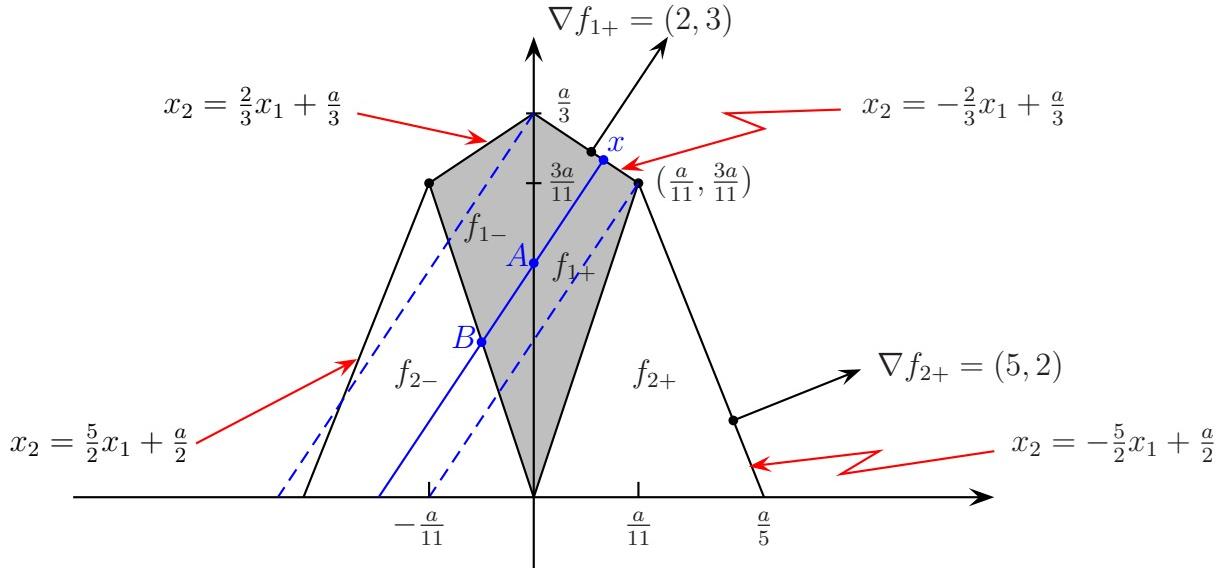
Consider the exceptional set $N = \cup_{i \neq j}\{f_i = f_j = f\}$, whose intersection with the upper half-plane $x_2 \geq 0$ consists of the three lines $x_1 = 0$, $x_2 = \pm 3x_1$. Then for $x \notin N$ the gradient $\nabla f(x)$ is unique. We will generate a sequence x^j of iterates which never meets N, so that $\phi^\sharp(y, x) = f(x) + \nabla f(x)^\top (y - x)$ with $\nabla f(x) \in \{\pm(2, 3), \pm(5, 2)\}$ at all iterates

x^j. It will turn out that serious iterates x^j never leave the dragon area, only trial points may.

Assume that our current iterate x has $f(x) = a$ and is situated on the right upper part of the a-dragon, shown as the blue x in the figure. That means

$$x = (x_1, -\tfrac{2}{3}x_1 + \tfrac{a}{3}), \quad f(x) = a, \quad 0 < x_1 \le \tfrac{a}{11}.$$

Then $\phi^\sharp(y, x) = f_{1+}(y) = 2y_1 + 3y_2$. If the current trust-region radius is $R = \sqrt{13}r$, then the solution of (2) is $y = x + r(-2, -3) = (x_1 - 2r, -\tfrac{2}{3}x_1 + \tfrac{a}{3} - 3r)$. If we follow the point y as a function of r along the steepest descent line shown in blue, we will reach the points A, B in increasing order at $0 < r_A < r_B$. Here A is the intersection of the steepest descent line with the x_2 axis, reached at $r_A = x_1/2$. The point B is when the ray meets the boundary of the a-dragon, which is the line $x_2 = -3x_1$ on the left, reached at

$$r_B = \tfrac{7}{27}x_1 + \tfrac{a}{27}.$$

We have $f(A) = f_{1+}(A) = a - \tfrac{17}{4}x_1$ and $f(B) = f_{1-}(B) = -\tfrac{143}{27}x_1 + \tfrac{22}{2}a$, and from here on f increases along the ray. The test quotient ρ for trial points y of this form behaves as follows

$$\rho = \frac{f(x_a) - f(y)}{f(x_a) - \phi^\sharp(y, x_a)} = \begin{cases} 1 & \text{if } 0 < r \le r_A \\ \frac{4x_1 + 5r}{13r} & \text{if } r_A \le r \le r_B \\ \frac{a - 12r + 19x_1}{39r} & \text{if } r_B \le r < \infty \end{cases}$$

The quotient is therefore constant on $[0, r_A]$, and decreasing on $[r_A, \infty)$. If we trace the quotient at the point B as a function of x_1, we see that $\rho = \tfrac{5}{13}$ at $x_1 = 0$, and $\rho = \tfrac{198}{234}$ at $x_1 = \tfrac{a}{11}$. That means if we take the Armijo constant as $\gamma \in (\tfrac{198}{234}, 1)$, then none of the points in $[B, \infty)$ is accepted, whatever $x_1 \in (0, \tfrac{a}{11}]$. Let the value r where the quotient ρ equals γ be called r_γ. Then $r_A < r_\gamma < r_B$, and we have $r_\gamma = \tfrac{4x_1}{13\gamma - 5}$.

Let us for simplicity put $\Gamma = 1$. That means good steps where the trust-region radius is doubled are exactly those in $(x, A]$, that is, $0 < r \le r_A$. Such a step is immediately accepted, and we stay on the right upper half of the a^+-dragon, where $a^+ < a$, except for the point A, which we will exclude later. We find for $0 < r < r_A = x_1/2$:

$$a^+ = a - 13r > 0, \quad x^+ = (x_1 - 2r, -\tfrac{2}{3}x_1 + \tfrac{a}{3} - 3r) = (x_1^+, -\tfrac{2}{3}x_1^+ + \tfrac{a^+}{3}).$$

Note that $a = a^+$ for the limiting case $x_1 = 0$, and $a^+ = \tfrac{9}{22}a$ for the limiting case $x_1 = \tfrac{a}{11}$. According to step 8 of the algorithm the trust-region radius is doubled ($R^+ = 2R$) for $0 < r < r_A$, because $\rho = 1 \ge \Gamma = 1$.

The second case is when from the current x with $f(x) = a$ a step with $R = \sqrt{13}r$ and $r \in (r_A, r_\gamma)$ is taken. Then we end up on the left hand side of the dragon with the new situation

$$x^+ = (x_1 - 2r, -\tfrac{2}{3}x_1 + \tfrac{a}{3} - 3r), \quad f(x^+) = f_{1-}(x^+) = -4x_1 + a - 5r = a^+.$$

By symmetry, this case is analogous to the initial situation, the model at x^+ now being f_{1-}. We are now on the upper left side of the smaller a^+-dragon. Since $\gamma \le \rho < \Gamma$, the trust-region radius remains unchanged.

The third case is when $r \in [r_\gamma, \infty)$. Here the step is rejected, and the trust-region radius is halved, until a value $r < r_\gamma$ is reached.

Since $\phi^\sharp$ is used, no cutting planes are taken, and we follow the classical trust-region method. In consequence, the serious iterates $x, x^+, x^{++}, \dots$ stay in the dragons $a, a^+, a^{++}, \dots$ and converge to the origin, which is not a critical point of f. Note that we have to assure that none of the trial points y lies precisely on the x_2-axis. Now it is clear that for a given starting point x the method has a countable number of possible trial steps y^k, and we can choose the initial $x_1 \in (0, \tfrac{a}{11}]$ such that the x_2-axis is avoided, for instance, by taking an

irrational initial value. Alternatively, in the case where y^k hits the x_2-axis, we might use rule (3) to change it slightly to a z^k, which is not on the axis. In both cases the method will never leave the dragon area, hence convergence based on the Cauchy point fails.

6. Parametric robustness

We consider an LFT plant [48] with real parametric uncertainties $\mathcal{F}_u(P, \Delta)$, where

$$(18) \qquad P(s): \begin{cases} \dot{x} &= Ax &+& B_p p &+& B_w w \\ q &= C_q x &+& D_{qp} p &+& D_{qw} w \\ z &= C_z x &+& D_{zp} p &+& D_{zw} w \end{cases}$$

and $x \in \mathbb{R}^{n_x}$ is the state, $w \in \mathbb{R}^{m_1}$ the vector of exogenous inputs, and $z \in \mathbb{R}^{p_1}$ the regulated output. The uncertainty channel is defined as $p = \Delta q$, where the uncertain matrix Δ is without loss assumed to have the block-diagonal form

$$(19) \qquad \Delta = \operatorname{diag}\left[\delta_1 I_{r_1}, \dots, \delta_m I_{r_m}\right]$$

with $\delta_1, \dots, \delta_m$ representing real uncertain parameters, and r_i giving the number of repetitions of δ_i. We write $\delta = (\delta_1, \dots, \delta_m)$ and assume without loss that $\delta = 0$ represents the nominal parameter value. Moreover, we consider $\delta \in \mathbb{R}^m$ in one-to-one correspondence with the matrix Δ in (19).

6.1. Worst case H_∞-performance over a parameter set.
Our first problem concerns analysis of the performance of a system (18) subject to parametric uncertainty. In order to analyze the robustness of (18) we compute the worst-case H_∞ performance of the channel $w \to z$ over a given uncertain parameter range normalized to $\mathbf{\Delta} = [-1, 1]^m$. In other words, we compute

$$(20) \qquad h^* = \max\{\|T_{wz}(\delta)\|_\infty : \delta \in \mathbf{\Delta}\},$$

where $T_{wz}(\delta)$ is the transfer function $z(s) = \mathcal{F}_u(P(s), \Delta)w(s)$, or more explicitly,

$$z(s) = \left[P_{22}(s) + P_{21}(s)\Delta(I - P_{11}(s)\Delta)^{-1}P_{12}(s)\right]w(s).$$

The significance of (20) is that computing a critical parameter value $\delta^* \in \mathbf{\Delta}$ which degrades the H_∞-performance of (18) may be an important domino in assessing the properties of a controlled system (18). We refer to [3] where this is exploited in parametric robust synthesis.

Solving (20) leads to a program of the form (1) if we write (20) as minimization of $h_-(\delta) = -\|T_{wz}(\delta)\|_\infty$ over the convex $\mathbf{\Delta}$. The specific form of $\mathbf{\Delta}$ strongly suggest the use of the maximum norm $|\delta|_\infty = \max\{|\delta_1|, \dots, |\delta_m|\}$ to define trust-regions. Moreover, we will use the standard model $\phi^\sharp$ of $h_-(\delta) = -\|T_{wz}(\delta)\|_\infty$, as is justified by the following

Lemma 6. *Let $D = \{\delta : T_{zw}(\delta)$ is internally stable$\}$. Then $h_- : \delta \mapsto -\|T_{zw}(\delta)\|_\infty$ is upper-C^1 on D.*

Proof. It suffices to prove that $h_+ : \delta \mapsto \|T_{wz}(\delta)\|_\infty$ is lower C^1. To prove this, recall that the maximum singular value has the variational representation

$$\overline{\sigma}(G) = \sup_{\|u\|=1} \sup_{\|v\|=1} \left|u^T G v\right|.$$

Now observe that $z \mapsto |z|$, being convex, is lower-C^1 as a mapping $\mathbb{R}^2 \to \mathbb{R}$, so we may write it as

$$|z| = \sup_{l \in \mathbb{L}} \Psi(z, l)$$

for Ψ jointly of class C^1 and a suitable compact set $\mathbb{L}$. Then

$$
(21) \qquad h_+(\delta) = \sup_{j\omega \in \mathbb{S}^1} \sup_{\|u\|=1} \sup_{\|v\|=1} \sup_{l \in \mathbb{L}} \Psi\left(u^T T_{zw}(\delta, j\omega)v, l\right),
$$

where $\mathbb{S}^1 = \{j\omega : \omega \in \mathbb{R} \cup \{\infty\}\}$ is homeomorphic with the 1-sphere. This is a representation of the form (17) for h_+, where the compact space is $\mathbb{K} := \mathbb{S}^1 \times \{u : \|u\| = 1\} \times \{v : \|v\| = 1\} \times \mathbb{L}$, F is $F(\delta, j\omega, u, v, l) := \Psi\left(u^T T_{zw}(\delta, j\omega)v, l\right)$ and $y = (j\omega, u, v, l)$. $\qquad \square$

Theorem 3 (Worst-case H_∞ norm on $\mathbf{\Delta}$). *Let $\delta^j \in \mathbf{\Delta}$ be the sequence generated by the standard trust-region algorithm applied to program (20) based on the standard model of h_-. Then the δ^j converge to a critical point δ^* of (20).*

Proof. By Lemma 5 Algorithm 1 coincides with a classical first-order trust-region algorithm, with convergence in the sense of subsequences. Convergence to a single critical point then follows by observing that h_- satisfies a Łojasiewicz inequality. $\qquad \square$

6.2. Robust stability over a parameter set.
In our second problem we wish to check whether the uncertain system (18) is robustly stable over the uncertain parameter set $\mathbf{\Delta} = [-1, 1]^m$. This can be tested by maximizing the spectral abscissa over $\mathbf{\Delta}$:

$$
(22) \qquad \alpha^* = \max\{\alpha\left(A(\delta)\right) : \delta \in \mathbf{\Delta}\},
$$

where $A(\delta)$ is the closed-loop system matrix

$$
(23) \qquad A(\delta) = A + B_p \Delta \left(I - D_{qp}\Delta\right)^{-1} C_q,
$$

and where the spectral abscissa of $A \in \mathbb{R}^{n \times n}$ is $\alpha(A) = \max\{\text{Re}(\lambda) : \lambda \text{ eigenvalue of } A\}$. The decision is now as follows. As soon as $\alpha^* \geq 0$, the solution δ^* of (22) represents a destabilizing choice of the parameters, and this may be valuable information in practice, see [3]. On the other hand, if the global maximum has value $\alpha^* < 0$, then a certificate for robust stability over $\delta \in \mathbf{\Delta}$ is obtained.

Global maximization of (22) is known to be NP-hard [34, 10], so it is interesting to use a local optimization method to compute good lower bounds. This can be achieved by algorithm 1, because (22) is clearly of the form (1) if maximization of α is replaced by minimization of $-\alpha$ over $\mathbf{\Delta}$. In our experiment additional speed is gained by adapting the trust-region norm $|\delta|_\infty = \max\{|\delta_1|, \ldots, |\delta_m|\}$ to the special form $\mathbf{\Delta} = [-1, 1]^m$ of the set C, and the standard model $\phi^\sharp$ of $a_-(\delta) = -\alpha(A(\delta))$ is used. With these arrangements the method converges fast and reliably to a local optimum, which in the majority of cases can be certified *a posteriori* as a global one.

In order to justify the use of the standard model in Algorithm 1 we have to show that a_- is upper-C^1, or at least that its standard model is strict. Here the situation is more delicate than in section 6.1. We start by observing the following

Lemma 7. *Suppose all active eigenvalues of $A(\delta)$ at δ are semi-simple. Then $a_-(\delta) = -\alpha\left(A(\delta)\right)$ is Clarke subdifferentiable in a neighborhood of δ.*

Proof. This follows from [11]. A very concise proof that semi-simple eigenvalue functions are locally Lipschitz could also be found in [27]. $\qquad \square$

That $a_\pm(\delta) = \pm\alpha(A(\delta))$ may fail to be locally Lipschitz was first observed in [11]. This may lead to difficulties when a_+ is minimized. In contrast, in our numerical testing it is $a_-(\delta) = -\alpha\left(A(\delta)\right)$ which is minimized, and this behaves consistently like an upper-C^1 function. Theoretically we expect a_- to have a strict standard model if all active eigenvalues of $A(\delta^*)$ are semi-simple. An argument indicating that its standard model is at least directionally strict is given in [3, V.C]. See [29] for more information on $a_\pm$.

Theorem 4 (Worst-case spectral abscissa on $\mathbf{\Delta}$). *Let $\delta^j \in \mathbf{\Delta}$ be the sequence generated by Algorithm 1 for program (22), where the standard model $\phi^\sharp$ of a_- is used. Suppose every accumulation point δ^* of the sequence δ^j is simple. Then the sequence δ^j converges to a critical point of (22).*

Proof. We apply Theorem 1 to get convergence in the sense of subsequences. $\qquad\square$

6.3. Distance to instability. Our third problem is related to the above and concerns computation of the structured distance to instability of (18). Suppose the matrix A in (18) is nominally stable, i.e., $A(\delta)$ is stable at the nominal $\delta = 0$. Then the structured distance to instability is defined as

$$(24) \qquad d^* = \max\{d > 0 : A(\delta) \text{ stable for all } |\delta|_\infty < d\},$$

where $A(\delta)$ is given by (23), and $|\delta|_\infty = \max\{|\delta_1|, \ldots, |\delta_m|\}$. Equivalently, we may consider the following constrained optimization program

$$(25) \qquad \begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -t \leq \delta_i \leq t \\ & \alpha\left(A(\delta)\right) \geq 0 \end{array}$$

with decision variable $x = (t, \delta) \in \mathbb{R}^{m+1}$. Introducing the convex set $C = \{(t, \delta) : -t \leq \delta_i \leq t, i = 1, \ldots, m\}$, this can be transformed to program (1) if we minimize an exact penalty objective $f(x) = t + c \max\{0, -\alpha\left(A(\delta)\right)\}$ with a penalty constant $c > 0$ over C.

It is clear that the objective of f has essentially the same properties as a_-. It suffices to argue that $\partial \max\{0, -\alpha(A(\delta))\} = \mathrm{co}\{0\} \cup \partial a_-(\delta)$ at points δ where a_- is locally Lipschitz and $a_-(\delta) = 0$. Indeed, the inclusion $\subset$ holds in general. For the reverse inclusion it suffices to observe that $0 \in \partial \max\{0, -\alpha(A(\delta))\}$ for those δ where $a_-(\delta) = 0$. This is clear, because 0 is a minorant of this max function. We may then use the following

Lemma 8. *Suppose $f = \max\{f_1, f_2\}$ and f_i has a strict model ϕ_i. Then $\phi = \max\{\phi_1, \phi_2\}$ is a strict model of f at those x where $\partial f(x) = \mathrm{co}\left(\partial f_1(x) \cup \partial f_2(x)\right)$.*

Proof. In fact, the only axiom which does not follow immediately is (M_1). We only know $\partial_1 \phi_i(x, x) \subset \partial f_i(x)$, so $\partial_1 \phi(x, x) = \mathrm{co}\left(\partial_1 \phi_1(x, x) \cup \partial_1 \phi_2(x, x)\right) \subset \mathrm{co}\left(\partial f_1(x) \cup \partial f_2(x)\right)$. For those x where the maximum rule is exact, this implies indeed $\partial_1 \phi(x, x) \subset \partial f(x)$. $\qquad\square$

This means that we can use the model $\phi(\delta', t', \delta, t) = t' + c \max\{0, \phi^\sharp(\delta', \delta)\}$ in Algorithm 1 to solve (25), naturally with the same proviso as in section 6.2, where we need the standard model $\phi^\sharp$ of a_- to be strict.

7. Experiments

In this part experiments with algorithm 1 applied to programs (20), (22) and (24) are reported.

7.1. Worst-case H_∞-norm. We apply algorithm 1 to program (20). Table 1 shows the result for 27 benchmark systems, where n is the number of states, and column 4 gives the uncertain structure $[r_1 \ldots r_m]$ according to (19). An expression like $1^3 3^1 1^1$ corresponds to $[r_1\, r_2\, r_3\, r_4\, r_5] = [1\, 1\, 1\, 3\, 1]$. The values achieved by algorithm 1 are h^* in column 6, obtained in t^* seconds CPU. To certify h^* we use the function WCGAIN of [49], which is a branch-and-bound method tailored to program (20). WCGAIN computes a lower and an upper bound $\underline{h}, \overline{h}$ shown in columns 5,7 within t_{wc} seconds. It also provides a $\underline{\delta} \in \mathbf{\Delta}$ realizing the lower bound.

The results in table 1 show that h^* is certified by WCGAIN in the majority of cases 1-5,7-9,11-13,16,17. Case 15 leaves a doubt, while cases 6,10,14,24 are failures of WCGAIN.

♯	**Benchmark**	n	**Structure**	$\underline{h}$	h^*	$\overline{h}$	t^*	$\overline{h}/h^*$	t_{wc}/t^*
1	Beam1	11	$1^3 3^1 1^1$	1.70	1.71	1.70	1.02	0.99	13.29
2	Beam2	11	$1^3 3^1 1^1$	1.29	1.29	1.29	0.36	1	32.68
3	DC motor 1	7	$1^1 2^2$	0.72	0.72	0.72	0.51	1.01	14.49
4	DC motor 2	7	$1^1 2^2$	0.50	0.50	0.50	0.13	1	45.02
5	DVD driver 1	10	$1^1 3^3 1^1 3^1$	45.45	45.45	45.46	0.23	1	189.31
6	Four-disk system 1	16	$1^1 3^5 1^4$	3.50	4.56	3.50	0.44	0.77	343.35
7	Four-disk system 2	16	$1^1 3^5 1^4$	0.69	0.68	0.69	0.34	1.01	558.03
8	Four-tank system 1	12	1^4	5.60	5.60	5.60	0.32	1	5.72
9	Four-tank system 2	12	1^4	5.60	5.57	5.60	0.29	1	7.32
10	Hard disk driver 1	22	$1^3 2^4 1^4$	243.9	7526.6	Inf	0.96	Inf	73.10
11	Hard disk driver 2	22	$1^3 2^4 1^4$	0.03	0.03	0.03	0.20	1.12	314.92
12	Hydraulic servo 1	9	1^9	1.17	1.17	1.17	0.34	1	10.94
13	Hydraulic servo 2	9	1^9	0.7	0.70	0.7	0.33	1.01	11.69
14	Mass-spring 1	8	1^2	3.71	6.19	3.71	0.31	0.60	3.54
15	Mass-spring 2	8	1^2	6.84	6.84	7.16	0.13	1.05	7.05
16	Missile 1	35	$1^3 6^3$	5.12	5.15	5.12	0.46	0.99	272.54
17	Missile 2	35	$1^3 6^3$	1.83	1.82	1.83	0.22	1	1183.5
18	Filter 1	8	1^1	4.86	4.86	4.86	0.32	1	3.41
19	Filter 2	3	1^1	2.63	2.64	2.63	0.27	1	4.06
20	Filter-Kim 1	3	1^2	2.95	2.96	2.95	0.24	1	3.4
21	Filter-Kim 2	3	1^2	2.79	2.79	2.79	0.07	1	12.95
22	Satellite 1	11	$1^1 6^1 1^1$	0.16	0.17	0.16	0.33	1	86.17
23	Satellite 2	11	$1^1 6^1 1^1$	0.15	0.15	0.15	0.70	1	41.09
24	Mass-spring-damper 1	13	1^1	7.63	8.85	7.63	0.21	0.86	4.88
25	Mass-spring-damper 2	13	1^1	1.65	1.65	1.65	0.08	1	13.70
26	Robust Toy 1	3	$1^1 2^1$	0.12	0.12	0.12	0.56	1	4.24
27	Robust Toy 2	3	$1^2 2^2 3^1$	20.85	21.70	20.91	0.24	0.96	29.19

TABLE 1. Benchmarks for worst-case H_∞-norm on $\boldsymbol{\Delta}$

On average algorithm 1 was 121-times faster than `WCGAIN`. The fact that both methods are in good agreement can be understood as an endorsement of our approach.

7.2. Robust stability over $\boldsymbol{\Delta}$. In our second test algorithm 1 is applied to program (22). We have used a bench of 32 cases gathered in Table 2, and algorithm 1 converges to the value α^* in t^* seconds. To certify α^* we have implemented algorithm 2, known as integral global optimization, or as the Zheng-method (ZM), based on [47]. Here μ is any

---

Algorithm 2. Zheng-method for global optimization $\alpha^* = \max_{x \in \boldsymbol{\Delta}} f(x)$

▷ **Step 1** (`Initialize`). Choose initial $\alpha < \alpha^*$.

▷ **Step 2** (`Iterate`). Compute $\alpha^+ = \dfrac{\int_{[f \geq \alpha]} f(x)\, d\mu(x)}{\mu[f \geq \alpha]}$.

▷ **Step 3** (`Stopping`). If progress of α^+ over α is marginal, stop, otherwise update α by α^+ and loop on with step 2.

---

continuous finite Borel measure on $\boldsymbol{\Delta}$. Numerical implementations use Monte-Carlo to compute the integral, and we refer to [47] for details. Our numerical tests are performed with $2000 \cdot m$ initial samples, and stopping criterion variance $= 10^{-7}$; cf. [47] for details. The result obtained by ZM are α_{ZM} obtained in t_{ZM} seconds CPU.

♯	Benchmark	n	Structure	α^*	α_{ZM}	t^*	t_{ZM}
28	Beam3	11	$1^33^11^1$	-1.2e-7	-1.2e-7	0.19	32.70
29	Beam4	11	$1^33^11^1$	-1.7e-7	-1.7e-7	0.04	33.00
30	Dashpot system 1	17	1^6	0.0186	0.0185	0.23	90.25
31	Dashpot system 2	17	1^6	-1.0e-6	-1.0e-6	0.39	39.63
32	Dashpot system 3	17	1^6	-1.6e-6	-1.6e-6	0.08	39.70
33	DC motor 3	7	1^12^2	-0.0010	-0.0010	0.02	20.63
34	DC motor 4	7	1^12^2	-0.0010	-0.0010	0.02	20.74
35	DVD driver 2	10	$1^13^31^13^1$	-0.0165	-0.0165	0.04	49.29
36	Four disk system 3	16	$1^13^51^4$	0.0089	0.0088	0.10	159.61
37	Four disk system 4	16	$1^13^51^4$	-7.5e-7	-7.5e-7	0.29	73.86
38	Four disk system 5	16	$1^13^51^4$	-7.5e-7	-7.5e-7	0.29	74.36
39	Four tank system 3	12	1^4	-6.0e-6	-6.0e-6	0.17	25.81
40	Four tank system 4	12	1^4	-6.0e-6	-6.0e-6	0.02	26.20
41	Hard disk driver 3	22	$1^32^41^4$	266.70	266.70	0.09	297.21
42	Hard disk driver 4	22	$1^32^41^4$	-1.6026	-1.6026	0.06	80.40
43	Hydraulic servo 3	9	1^9	-0.3000	-0.3000	0.04	51.41
44	Hydraulic servo 4	9	1^9	-0.3000	-0.3000	0.02	50.95
45	Mass-spring 3	8	1^2	-0.0054	-0.0054	0.01	31.59
46	Mass-spring 4	8	1^2	-0.0368	-0.0370	0.01	16.94
47	Missile 3	35	1^36^3	22.6302	22.1682	0.07	104.18
48	Missile 4	35	1^36^3	-0.5000	-0.5000	0.07	51.78
49	Missile 5	35	1^36^3	-0.5000	-0.5000	0.07	52.24
50	Filter 3	8	1^1	-0.0148	-0.0148	0.06	7.05
51	Filter 4	8	1^1	-0.0148	-0.0148	0.02	6.89
52	Filter-Kim 3	3	1^2	-0.2500	-0.2500	0.01	12.83
53	Filter-Kim 4	3	1^2	-0.2500	-0.2500	0.01	12.90
54	Satellite 3	11	$1^16^11^1$	3.9e-5	3.9e-5	0.02	44.02
55	Satellite 4	11	$1^16^11^1$	-0.0269	-0.0269	0.02	26.02
56	Satellite 5	11	$1^16^11^1$	-0.0268	-0.0268	0.02	26.08
57	Mass-spring-damper 3	13	1^1	0.2022	0.2022	0.01	8.30
58	Mass-spring-damper 4	13	1^1	-0.1000	-0.1000	0.01	6.91
59	Mass-spring-damper 5	13	1^1	-0.1000	-0.1000	0.01	6.94

TABLE 2. Benchmarks for worst-case spectral abscissa (22).

A favorable feature of ZM is that it can be initialized with the lower bound α^*, and this leads to a significant speedup. Altogether ZM and algorithm 1 are in very good agreement on the test bench, which we consider an argument in favor of our approach.

7.3. Distance to instability. In this last part we apply Algorithm 1 to (24) using the test bench of Table 3, which can be found in [18]. The distance computed by Algorithm 1 is d^* in column 2 of Table 3. We certify d^* using ZM [47] and by comparing to the local method of [18].

To begin with, ZM is used in the following way. For a given d^* and a confidence level $\gamma = 0.05$ we compute

$$(26) \qquad \underline{\alpha} = \max\{\alpha(A(\delta)) : \delta \in (1-\gamma)d^*\boldsymbol{\Delta}\}$$

and

$$(27) \qquad \overline{\alpha} = \max\{\alpha(A(\delta)) : \delta \in (1+\gamma)d^*\boldsymbol{\Delta}\}.$$

♯	Benchmark	n	Structure	d^*	d_{F}/d^*	$\mathbf{D}_{\mathrm{ZM}}$	t^*	t_{ZM}
60	Academic example	5	1^1	0.79	1	$\surd$	0.15	7.3
61	Academic example	4	1^3	3.41	1	$\surd$	0.13	23.9
62	Academic example	4	2^2	0.58	1	$\surd$	0.15	97.4
63	Inverted pendulum	4	1^3	0.84	1	$\surd$	0.22	24.7
64	DC motor	4	$1^3 2^1\,1^1$	1.25	1	$\surd$	0.19	37.7
65	Bus steering system	9	$2^1 3^1$	1.32	0.99	$\surd$	0.37	13.8
66	Satellite	9	$2^1 1^2$	1.01	0.99	$\surd$	0.3	20.2
67	Bank-to-turn missile	6	1^4	0.60	0.99	$\surd$	0.17	167.7
68	Aeronautical vehicle	8	1^4	0.61	0.99	$\surd$	0.19	38.9
69	Four-tank system	10	1^4	6.67	0.99	$\surd$	0.27	24.9
70	Re-entry vehicle	6	$3^1 2^1 3^1$	6.20	1	$\surd$	0.44	21.8
71	Missile	14	1^4	7.99	1	$\surd$	0.25	24.9
72	Cassini spacecraft	17	1^4	0.06	1	$\surd$	0.13	25.1
73	Mass-spring-damper	7	1^6	1.17	1	$\surd$	0.17	2536.3
74	Spark ignition engine	4	1^7	1.22	0.99	$\surd$	0.41	42.8
75	Hydraulic servo system	8	1^8	1.50	0.99	$\surd$	0.41	62.8
76	Academic example	41	$2^1 1^3$	1.18	0.99	$\surd$	0.57	36.5
77	Drive-by-wire vehicle	4	$1^2 2^7$	1	0.99	$\surd$	0.96	97.0
78	Re-entry vehicle	7	$1^3 6^1 4^1$	1.02	0.98	$\surd$	0.42	132.4
79	Space shuttle	34	1^9	0.79	0.99	$\surd$	0.8	60.9
80	Rigid aircraft	9	1^{14}	5.42	1	$\surd$	0.54	252.5
81	Fighter aircraft	10	$3^1 15^1 6^2 1^1 1^1$	0.59	0.99	$\surd$	1.31	171.3
82	Flexible aircraft	46	1^{20}	0.22	0.99	$\surd$	1.26	180.3
83	Telescope mockup	70	1^{20}	0.02	0.99	$\surd$	1.37	274.8
84	Hard disk drive	29	$1^8 2^4 1^1 1$	0.82	1	$\surd$	2.87	202.1
85	Launcher	30	$1^2 2^2 1^2 3^1 6^1 1^{12} 2^8$	1.16	0.99	$\surd$	4.08	271.2
86	Helicopter	12	30^4	0.08	0.99	$\surd$	0.85	70.7
87	Biochemical network	7	39^{13}	0.00	1	failed	36.76	-

TABLE 3. Benchmarks for distance to instability (24), available in [50].

If $\underline{\alpha} < 0$ and $\overline{\alpha} > 0$ then d^* is certified by ZM with that confidence level γ. This happens in all cases except 87, where ZM failed due to the large size.

We also compared d^* to the result d_{F} of the technique [18], which is a sophisticated tool tailored to problem (24). Column 6 of table 3 shows perfect agreement on the bench from [18]. Given the highly dedicated character of [18], this can be understood as an endorsement of our optimization-based approach.

Conclusion

We have presented a bundle trust-region method for nonsmooth, nonconvex minimization, where cutting planes are tangents to a convex local model $\phi(\cdot, x)$ of f, and where a trust-region strategy replaces the proximity control mechanism. Global convergence of our method was proved under natural hypotheses.

By way of an example we demonstrated that the standard approach in trust-region methods based on the Cauchy point fails for nonsmooth functions. We have identified a particular class $\mathscr{S}$ of nonsmooth functions, where the Cauchy point argument can be salvaged. Functions in $\mathscr{S}$, even when nonsmooth, can be minimized as if they were smooth. The class $\mathscr{S}$ must therefore be regarded as atypical in a nonsmooth optimization program, and indeed, nonsmooth convex functions are not in $\mathscr{S}$.

Algorithm 1 was validated numerically on a test bench of 87 problems in automatic control, where the versatility of algorithm 1 with regard to the choice of the norm was exploited. We were able to compute good quality lower bounds for three NP-hard optimization problems related to the analysis of parametric robustness in system theory. In the majority of cases, posterior application of a global optimization technique allowed us to certify these results as globally optimal.

References

[1] P.A. Absil, R. Mahony, B. Andrews. Convergence of the iterates of descent methods for analytic cost functions. *SIAM Journal on Optimization*, 16(2):531–547, 2005.

[2] S. Adly, H. Attouch, A. Cabot. Finite time stabilization of nonlinear oscillators subject to dry friction. Nonsmooth mechanics and analysis. *Adv. Mech. Math.* 12, Springer, New York, 2006, 289 – 304.

[3] P. Apkarian, M.N. Dao, D. Noll, Parametric robust structured control design. *IEEE Trans. Autom. Control*, to appear 2015.

[4] P. Apkarian, D. Noll. Nonsmooth H_∞ synthesis. *IEEE Trans. Automat. Control* 51(1) (2006), 71-86.

[5] P. Apkarian, D. Noll. Nonsmooth optimization for multidisk H_∞ synthesis. *Eur. J. Control* 12(3) (2006), 229-244.

[6] P. Apkarian, D. Noll, IQC analysis and synthesis via nonsmooth optimization. *Systems and Control Letters*, vol. 55, no. 12, p. 971 - 981.

[7] P. Apkarian, D. Noll, O. Prot. A proximity control algorithm to minimize non-smooth and nonconvex semi-infinite maximum eigenvalue functions. *Journal of Convex Analysis,* vol. 16, 2009, pp. 641 – 666.

[8] H. Attouch, J. Bolte, P. Redont, A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Journal Mathematics of Operations Research*, 35(2), 2010.

[9] J.M. Borwein, W.B. Moors. A chain rule for essentially strictly differentiable Lipschitz functions. *SIAM J. Optim.* 8 (1998), 300-308.

[10] R. D. Braatz and P. M. Young and J. C. Doyle and M. Morari. Computational complexity of μ calculation. *IEEE Transactions on Automatic Control*, 39, 1994, 1000–1002.

[11] J. V. Burke, M. L. Overton. Differential properties of the spectral abscissa and the spectral radius for analytic matrix-valued mappings. *Nonlinear Anal.* 23 (1994), no. 4, 467-488.

[12] F. H. Clarke. *Optimization and Nonsmooth Analysis.* John Wiley & Sons, Inc., New York, 1983.

[13] A.R. Conn, N.I.M. Gould, Ph.L. Toint. *Trust-region methods.* MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2000.

[14] Daniilidis, A., Georgiev, P.: Approximate convexity and submonotonicity. *J. Math. Anal. Appl.* 291, 117-144 (2004).

[15] M.N. Dao. Bundle method for nonconvex nonsmooth constrained optimization. *Journal of Convex Analysis*, to appear.

[16] M.N. Dao, J. Gwinner, D. Noll, N. Ovcharova. Nonconvex bundle method with application to a delamination problem. `arXiv:1401.6807v1 [math.OC] 27 Jan 2014`.

[17] J.E. Dennis, S.B. Li, R.A. Tapia. A unified approach to global convergence of trust-region methods for nonsmooth optimization. *Math. Programming* 68 (1995), 319–346.

[18] A. Fabrizi, C. Roos, J.M. Biannic. A detailed comparative analysis of lower bound algorithms. *European Control Conference* 2014, Jun 2014, Strasbourg, France.

[19] A. Fuduli, M. Gaudioso and G. Giallombardo. A DC piecewise affine model and a bundling technique in nonconvex nonsmooth optimization. *Optimization Method and Software*, vol. 19, 2004, p. 89 - 102.

[20] C. Helmberg, K.C. Kiwiel. A spectral bundle method with bounds. *Math. Programming*, vol. 93, 2002, p. 173 - 194.

[21] C. Helmberg, F. Oustry. Bundle methods to minimize the maximum eigenvalue function. Handbook of Semidefinite Programming. Theory, Algorithms and Applications. L. Vandenberghe, R. Saigal, H. Wolkowitz (eds.), vol. 27, 2000.

[22] C. Helmberg, F. Rendl. Spectral bundle method for semidefinite programming. *SIAM J. Optimization*, vol. 10, 2000, p. 673 - 696.

[23] Hiriart-Urruty, Lemaréchal. Convex Analysis and Minimization Algorithms, vol. I and II: Advanced Theory and Bundle Methods, vol. 306 of Grundlehren der mathematischen Wissenschaften, Springer Verlag, New York, Heidelberg, Berlin, 1993.

[24] C. Kiwiel, An aggregate subgradient method for nonsmooth convex minimization. *Math. Programming*, vol. 27, 1983, p. 320 - 341.

[25] D. Klatte, B. Kummer. *Nonsmooth Equations in Optimization. Regularity, Calculus, Methods and Applications.* Nonconvex Optimization and Applications. Kluwer Academic Publishers, 2002, vol. 60.

[26] Le Thi Hoai An, Huynh Van Ngai, Pham Dinh Tao, A. Ismael, F. Vaz, L. N. Vicente. Globally Convergent DC Trust-Region Methods. *Journal of Global Optimization*, 59 (2014) 209-225

[27] S. H. Lui, Pseudospectral mapping theorem II, *Electron. Trans. Numer. Anal.*, 38, 168–183, 2011.

[28] R. Mifflin. Semismooth and semiconvex functions in optimization. *SIAM J. Control and Optimization*, 15(6):1977, 959–972.

[29] J. Moro, J. V. Burke, M. L. Overton. On the Lidskii-Vishik- Lyusternik perturbation theory for eigenvalues of matrices with arbitrary Jordan structure. *SIAM J. Matrix Anal. Appl.*, 18(4) 1997, 793–817.

[30] Y. Nesterov. Smoothing technique and its applications in semidefinite optimization. *Math. Program.*, Ser. A 110 (2007), no. 2, 245-259.

[31] D. Noll. Cutting plane oracles to minimize non-smooth non-convex functions. *Set-Valued Var. Anal.* 18 (2010), no. 3-4, 531-568.

[32] D. Noll. Convergence of non-smooth descent methods using the Kurdyka-Łojasiewicz inequality. *J. Optim. Theory Appl.* 160 (2014), no. 2, 553-572.

[33] D. Noll, O. Prot, A. Rondepierre. A proximity control algorithm to minimize nonsmooth and non-convex functions. *Pac. J. Optim.* 4 (2008), no. 3, 571-604.

[34] S. Poljak, J. Rohn. Checking robust nonsingularity is NP-hard. *Math. Cont. Sig. Sys.* 6 (1993), 1–9.

[35] M.J.D. Powell, General algorithms for discrete nonlinear approximation calculations, Report DAMTP 1983/NA2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England (1983).

[36] M. Raous. Friction and adhesion. Nonsmooth Mechanics and Analysis, Chapter 9. Advances in Mechanics and Mathematics, vol. 12. (P. Alart, O. Maisonneuve, R.T. Rockafellar (eds.)).

[37] Rockafellar, R.T., Wets, R. J-B.: *Variational Analysis.* Springer Verlag (2004).

[38] A. Ruszczyński. Nonlinear Optimization. Princeton University Press, 2007.

[39] C. Sagastizábal, W. Hare. A redistributed proximal bundle method for nonconvex optimization. *SIAM J. Optim.* vol. 20, no. 5, 2010, pp. 2442 – 2473.

[40] C. Sagastizábal. Composite proximal bundle method. Preprint 2009.

[41] Y. Sartenaer. Armijo-type condition for the determination of a Generalized Cauchy Point in trust region algorithms using exact or inexact projections on convex constraints. *Belgian Journal of Operations Research, Statistics and Computer Science* 33(4): 61–75.

[42] H. Schramm. Eine Kombination von Bundle- und Trust-Region-Verfahren zur Lösung nicht-differenzierbarer Optimierungsprobleme. *Bayreuther Mathematische Schriften*, 30, Bayreuth 1989.

[43] H. Schramm, J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM J. Optim.* vol. 2, 1992, p. 121 - 152.

[44] J. E. Spingarn. Submonotone subdifferentials of Lipschitz functions. *Trans. Amer. Math. Soc.* 264 (1981), no. 1, 77-89.

[45] K. Zhou, J. C. Doyle, K. Glover. *Robust and Optimal Control.* Prentice Hall, New Jersey, 1996.

[46] J. Zowe. The BT-Algorithm for minimizing a nonsmooth functional subject to linear constraints. In Nonsmooth Optimization and Related Topics , F. H. Clarke, V. F. Demyanov, F. Gianessi (eds.), Plenum Press, 1989.

[47] Q. Zheng and D. Zhuang. Integral global minimization: algorithms, implementations, and numerical tests. *Journal of Global Optimization* 7 (1995), 421 – 454.

[48] K. Zhou, J. C. Doyle, K. Glover. *Robust and Optimal Control.* Prentice Hall, New Jersey, 1996.

[49] Robust Control Toolbox 5.0. MathWorks, Natick, MA, USA, Sept 2013.

[50] SMAC Toolbox, ONERA 2012-15, http://w3.onera.fr/smac